The Dissertation Committee for Pravesh Kumar Kothari
certifies that this is the approved version of the following dissertation:

Strong Lower Bounds on Generic Convex Relaxations

Committee:

Adam Klivans, Supervisor

Boaz Barak

Eric Price

Pradeep Ravikumar

David Zuckerman

Strong Lower Bounds on Generic Convex Relaxations

by

Pravesh Kumar Kothari, B. Tech.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

August 2016

To Arpita and my parents.

Acknowledgments

I am greatly indebted to my adviser Adam Klivans. From the beginning, Adam was consistently patient and encouraging and provided advice and guidance at every step. In the later years of my PhD, Adam gave me the freedom to travel and work with at various places and branch out into different research areas while continuing to provide valuable advice, help and expertise. This has been instrumental in doing the work presented in this thesis.

I am very thankful to my *pseudo*-adviser Boaz Barak who has been a constant guide in most of the work presented in this thesis. Working with Boaz and observing his enthusiasm in choosing and attacking problems has been a very inspiring experience - it almost appears as if failure is not an option when working with him! I am also grateful to him for the numerous philosophical discussions about everything in general and theoretical computer science in particular. They have been invaluable in shaping my aesthetics in research.

I am very grateful to Raghu Meka for advice and help at various points both in research and otherwise. I am also thankful to Prasad Raghavendra for inviting me to visit him in Berkeley in summer 2014 and then later, along with Raghu, in Fall 2015 where some of the work in this thesis began.

I am thankful to Madhu Sudan for valuable advice and the many elightening research discussions over the past couple years. Madhu introduced me to the exciting

area of communication under uncertainty and it was a great learning experience to work and think with him. Almost equally important is the tradition of beer after the Friday reading group seminar that has been a constant feature of the last two years in my PhD.

I am indebted to Prateek Jain, Deeparnab Chakrabarty, Nisheeth Vishnoi, Ramarathnam Venkatesan, Vitaly Feldman and Jan Vondrák for hosting me at Microsoft Research (to Bangalore, Cambridge, MA and the Mountain View labs in particular) and IBM Research Almaden for hosting me at various times! I am thankful to Ryan O'Donnell for hosting me at Carnegie Mellon in Fall 2012. Ryan is an incredible person to do research with and it's a challenge to match his energy and enthusiasm. I am thankful to Avi Wigderson and Ran Raz for hosting me at the IAS, to Sanjeev Arora and Assaf Naor for arranging a visit to Princeton University and to David Steurer for hosting me at Cornell and a fun week of research and coffee in Ithaca. Many thanks to my thesis committee members Boaz, Eric Price, Pradeep Ravikumar and David Zuckerman for their valuable comments on the proposal and the thesis and being flexible with all the scheduling of the proposal and defense. I am very grateful to friends and office mates for the many discussions on random topics and coffee breaks at odd hours. Special thanks to Lydia Griffith, Katie Dahm and all the other competent and helpful staff at UTCS.

Lastly, I'd like to thank my parents Sulabha and Omprakash Kothari and my sister Prerana, for their constant encouragement and support for the choices I made. I am especially thankful to my girlfriend Arpita who has been a constant companion and an unending source of encouragement throughout my PhD years. It is hard to imagine this work without her being around at every step to brighten up even the most unsuccessful days.

Strong Lower Bounds on Generic Convex Relaxations

Publication No. _____

Pravesh Kumar Kothari, Ph.D.
The University of Texas at Austin, 2016

Supervisor: Adam Klivans

Despite significant successes in understanding the hardness of computational problems based on standard assumptions such as $\boldsymbol{P \neq NP}$, there are important settings where the gap between what the best known algorithms achieve and what the best known hardness reductions can rule out is rather stark. This thesis aims at decreasing this gap by proving unconditional lower bounds on powerful algorithmic techniques based on linear (LP) and semi-definite programming (SDP) for *Planted Clique* and *Constraint Satisfaction* problems (CSPs).

Planted Clique is a central question in average case complexity where the goal is to find a clique of size ω planted in the Erdős-Rényi random graph $G(n, \frac{1}{2})$. While information theoretically, such a clique can be found whenever $\omega \gg 2\log(n)$, state-of-the-art polynomial time algorithms succeed only when $\omega = \Omega(\sqrt{n})$. In fact, the conjectured hardness of detecting planted cliques for $\omega \ll \sqrt{n}$ has been used as a starting point in many works [8, 5, 31, 2, 12, 3].

We show that algorithms with running time $n^{o(\log(n))}$ based on the sum-of-squares method cannot detect planted cliques when $\omega \ll \sqrt{n}$. *Sum-of-Squares* method [100, 90, 80]

is a general scheme based on SDP that extends natural LP relaxations and spectral methods, captures the best approximation algorithms for fundamental problems [54, 11, 9] and is the prime candidate for refuting the Unique Games Conjecture [76].

Further, we show that any sub-exponential algorithm based on the sum-of-squares method cannot outperform simply returning an assignment at random for any *pairwise independent* CSP - a large class of CSPs not known be NP-hard to approximate beyond the random assignment threshold. Finally, we show that sub-exponential size (number of inequalities/constraints) LP [107, 38] that encodes MAX-CSP as a linear optimization over any fixed polytope are no more powerful than the well-known Sherali-Adams LP of similar size. This immediately yields a sub-exponential size lower bound for any LP that obtains $> \frac{1}{2}$ approximation for MaxCut showing an exponential separation between linear and semidefinite programming.

Table of Contents

List of Figures

Chapter 1

Introduction

The broad goal of this work is to understand the complexity of *combinatorial optimization* problems. In combinatorial optimization problems, the algorithmic task is to optimize a specified cost function over some finite space of solutions. Concretely, we will use MaxCut and Clique as our running examples of such problems in this chapter.

MaxCut :*Given a graph, find the cut that includes the largest number of edges.*

Clique : *Given a graph, find the largest clique in it.*

The complexity of *exactly* solving such problems is by and large well understood. While there are polynomial time algorithms for some problems such as such as MinSpanningTree and MaxMatching, most problems turn out to be intractable under standard assumptions such as $\boldsymbol{P \neq NP}$ or the Exponential Time Hypothesis(ETH) [68]. As a result, the general idea of relaxing the notion of solution in order to gain tractability has been a topic of extensive research. We will focus on two such well-studied frameworks in this thesis.

Approximation Complexity. Approximation complexity studies the complexity of computing approximately optimal solutions to combinatorial optimization problems. Unlike the case of exact computation discussed before, in this regime, problems display interest-

ing variations in their computational complexity. Thus, while constant factor approximations are possible for problems such as constraint satisfaction and VERTEX COVER, for others such as SET COVER, one can only hope for approximation factors that grow with the input size [44] and for some other problems such as CLIQUE , obtaining any sub-linear approximation factor [62] turns out to require super polynomial time assuming $\boldsymbol{P \neq NP}$. Further, the gaps between the approximation guarantees that the best known efficient algorithms can provide and what the best known hardness results can rule out is rather stark for many natural problems. State of the art algorithms are typically based on linear/semidefinite programming relaxations followed by a rounding scheme. For example, the seminal work of Goemans and Williamson [54] used semidefinite programming to give an algorithm that returns a cut of size within 0.878 factor that of the optimal in the MAXCUT problem. On the other hand, the sophisticated machinery of Håstad [63] can be used to show that obtaining an approximation guarantee of 0.94 or better requires sub-exponential time under the ETH.

Average Case Complexity. Average case complexity (or more accurately, *typical* case complexity, see the survey [55]) studies the question of computing (approximately) optimal solutions for *typical* instances of the problem defined with respect to some underlying natural distribution on instances. One well-studied relaxation of CLIQUE in this paradigm is the PLANTEDCLIQUE problem: where given a Erdős-Rényi random graph with an added ω size clique, the task is to recover the added clique. While information theoretically, such a task is feasible whenever $\omega = \Omega(\log(n)$, state-of-the-art efficient algorithms [6] succeed only when $\omega = \Omega(\sqrt{n})$. Average case complexity is the most relevant notion

of complexity for problems in machine learning and statistical physics [31, 41] where problem instances are generally associated with a natural probability distribution. As above, however, this measure of complexity from being understood even for natural problems based on standard assumptions such as $\boldsymbol{P \neq NP}$ or ETH and typically, the best known algorithms are based on linear/semidefinite programming techniques.

Coping with *Lack* of NP Hardness. Given the above context the broad goal of this thesis is to prove lower bounds on strong, general purpose linear and semidefinite programming based algorithmic schemes in order to gain credible evidence of hardness and predict threshold values for relevant parameters such as the approximation factors/signal-to-noise-ratio where the computational complexity of natural problems potentially suffers a transition.

Seen differently, since linear/semidefinite programming based algorithms have been the most powerful tools used in algorithm design for combinatorial optimization, lower bounds for algorithmic schemes based on these techniques can be interpreted as revealing the structure in the instances that renders them hard for such algorithms.

Next, we state the results in this thesis and discuss the implications to the broad goals outlined above. The results in this thesis focus on the PlantedClique and *Constraint Satisfaction* problems (of which MaxCut above is an example).

1.1 Approximation Complexity of Constraint Satisfaction

We begin by revisiting the approximation complexity MaxCut problem. Simply returning a random cut in the graph achieves an approximation factor of 0.5 in expec-

tation. As we saw earlier, the semidefinite programming based algorithm of Goemans and Williamson [54] allows one to beat this guarantee substantially. How general is this phenomenon?

More generally, MaxCut is an instance of a class of natural computational problems known as *Constraint Satisfaction* that includes for example Max3Sat, Max3Xor and UniqueGames. *Constraint Satisfaction Problems* (CSP) are among the most natural computational problems, and yet the approximation complexity of CSPs is not fully understood.

Formally, a CSP is defined by a predicate $P : \{-1,1\}^k \to \{0,1\}$. An instance of the CSP, $\mathcal{I}$, is described by a collection of m k-tuples of literals (i.e. the variables or their negations) $C_1, C_2, \ldots, C_m$ on n Boolean variables $x_1, x_2, \ldots, x_n$. The algorithmic problem is to find an assignment $x = (x_1, x_2, \ldots, x_n) \in \{-1,1\}^n$ to maximize the number of satisfied constraints:

$$\mathrm{opt}(\mathcal{I}) = \max_{x\in\{-1,1\}^n} \sum_{i=1}^{m} P(C_i(x)) \equiv \max_{x\in\{-1,1\}^n} \mathcal{I}(x), \tag{1.1.1}$$

where we define $\mathcal{I}(x) = \sum_{i=1}^{m} P(C_i(x))$.

As in the case of MaxCut above, choosing x uniformly at random from $\{-1,1\}^n$ achieves an approximation factor of $\frac{|P^{-1}(1)|}{2^k}$ in expectation. Unlike the case of MaxCut , it turns out that there are CSPs where one can provably rule out efficient algorithms that beat the approximation guarantee of the random assignment assuming $\boldsymbol{P \neq NP}$ [63], such CSPs are said to be *approximation resistant.* We can also consider approximation resistance with respect to specific classes of algorithms where, in general, and ask whether efficient algorithms from a specific restricted class beats the approximation guarantee of

the random assignment.

1.1.1 Linear Programs and CSPs

As a concrete instance of the question above, we can ask:

Question 1. *Is there a linear program of size sub-exponential in n that approximates the* MaxCut *value in any graph on n vertices within a factor of* 0.51*?*

Since linear programming is P-complete under some notions of reduction, one can only hope to answer the question above that for linear programs that use "natural" encoding of the underlying combinatorial problem. Such linear programs were first formalized as *extended formulations* in the seminal work of Yannakakis [107] who showed that any *symmetric* linear program for that exactly solves TravelingSalesmanis of exponential size. Surprisingly, he showed this via reduction to MaximumMatching which does have a polynomial time algorithm. The framework was generalized to approximate complexity by Braun et. al. [35] who proved linear programming size lower bounds for approximating Clique and formally posed the question of extension complexity of approximating CSPs beyond the random assignment threshold.

Specifically, we consider a broad-class of linear programming relaxations for CSPs obtained by *linearizing* the objective function $\mathcal{I}(x)$. This framework was introduced in the work of [38] and generalizes the *extended formulation* framework of Yannakakis [107]. Formally, given a predicate P, and an integer D, we first need a linearization of the objective function $\mathcal{I}(x)$:

Definition 1.1.1 (Linearization)**.** A linearization for CSP with predicate P is defined by giving:

1. A vector $v_x \in \mathbb{R}^D$ for every $x \in \{-1,1\}^n$.
2. A vector $w_{\mathcal{I}} \in \mathbb{R}^D$ for every instance $\mathcal{I}$ of the CSP.
3. The vectors v_x and $v_{\mathcal{I}}$ satisfy the condition that for every assignment x and every instance $\mathcal{I}$, $\mathcal{I}(x) = \langle w_{\mathcal{I}}, v_x \rangle$.

Given a linearization as above, a polytope $\mathcal{P} \subseteq \mathbb{R}^{\mathcal{D}}$ such that $\{v_x : x \in \{-1,1\}^n\} \subseteq \mathcal{P}$ defines a relaxation for the CSP as follows. Given an instance, we look at the linear program

$$\mathrm{opt}_{\mathcal{P}}(\mathcal{I}) = \max_{y \in \mathcal{P}} \langle w_{\mathcal{I}}, y \rangle.$$

Clearly, $\mathrm{opt}(\mathcal{I}) \leqslant \mathrm{opt}_{\mathcal{P}}(\mathcal{I})$. The complexity of the relaxation is defined as the number of facets (or inequalities) needed to describe the polytope $\mathcal{P}$.

(c,s)-approximation. We say that $\mathcal{P}$ achieves an α-approximation factor if for every valid instance $\mathcal{I}$, $\mathrm{opt}_{\mathcal{I}} \geqslant \alpha\, \mathrm{opt}_{\mathcal{P}}(\mathcal{I})$. More generally, we say that $\mathcal{P}$ achieves a (c,s)-*approximation* if for all valid instances $\mathcal{I}$ such that $\mathrm{opt}(\mathcal{I}) \leqslant s$, $\mathrm{opt}_{\mathcal{P}}(\mathcal{I}) \leqslant c$.

In Chapter 2 of this thesis, we show that despite its apparent vast generality, when it comes to CSPs, general linear programs as above are only as powerful as those obtained from the Sherali-Adams hierarchy. These results are based on joint work with Raghu Meka and Prasad Raghavendra [78].

Theorem 1.1.2. *There exist constants $0 < h < H < 1$ such that the following holds. Consider a function $f : \mathbb{N} \to \mathbb{N}$. Suppose that the $f(n)$-round Sherali-Adams relaxation for a CSP*

cannot achieve a (c, s)*-approximation. Then, no LP relaxation of size* $n^{hf(n)}$ *can achieve a* (c, s)*-approximation for the CSP on* n^H *variables.*

In conjunction with known lower bounds for Sherali-Adams relaxations, the above theorem yields the following corollary:

Corollary 1.1.3. *For every* $\varepsilon > 0$*, there exist constants* $c_\varepsilon, c'_\varepsilon$ *such that the following holds. No LP relaxation of size less than* $2^{n^{c_\varepsilon}}$ *can achieve a* $(7/8 + \varepsilon)$*-factor approximation for MAX-3SAT. Similarly, no LP relaxation of size less than* $2^{n^{c'_\varepsilon}}$ *can achieve a* $(1/2 + \varepsilon)$*-factor approximation for MAX-CUT.*

Previously, Chan et. al. [38] invented an elegant technique to show, via a connection between arbitrary linear programming formulations as above and the Sherali Adams hierarchy that beating the random assignment for MaxCut requires $n^{\Omega(\log(n)/\log\log(n)}$ size linear programs. Our proof of Theorem 1.1.2 above is based on a general, tight relationship between the *non-negative degree* of a function and the *non-negative rank* of the pattern matrix of f.

One of the conclusions of the result above is an exponential separation between the power of linear and semidefinite programming for CSPs - while a simple SDP achieves 0.87 approximation for MaxCut , our result shows that no LP of even sub-exponential size achieves 0.51 approximation.

1.1.2 Sum-of-Squares Method and CSPs

We continue our investigation of approximation complexity of CSPs and in particular the phenomenon of approximation resistance. In the previous section we studied the

efficacy of linear programming for CSPs. As discussed several times before, semidefinite programming is known to do much better for a number of CSPs and in fact is the technique of choice for beating random assignment based guarantees. On the other hand, as previously mentioned, there are CSPs where the analog of Goemans-Williamson semidefinite program doesn't succeed in beating the random assignment.

This brings us to the next issue of our interest: *Can we obtain a structural understanding of predicates P for which it is possible to beat the random assignment efficiently?*

Under Khot's *Unique Games Conjecture* [76] much is known regarding approximation resistance of CSPs. In particular Austrin and Mossell [15] showed if the UGC is true, then, for every predicate $P : \{0,1\}^k \to \{0,1\}$, if there exists a *pairwise independent* distribution μ over $P^{-1}(1)$ (i.e., a distribution μ such that for every $i \neq j \in [k]$, the marginal $\mu_i\mu_j$ is the uniform distribution over $\{0,1\}^2$), then P is approximation resistant. Austrin and Håstad [14] used this to establish (under the UGC) fairly tight bounds on the threshold at which a random predicate of a particular density becomes approximation resistant. However, there is no consensus whether the UGC is true. Assuming only $\boldsymbol{P \neq NP}$, the best known bound is by Chan [37] who showed that a predicate is approximation resistant if it contains a distribution μ as above satisfying the additional condition that it is uniform over a subspace $V \subseteq GF(2)^k$. This algebraic structure is a fairly strong condition. In particular if we choose $P : \{0,1\}^k \to \{0,1\}$ to be a random predicate conditioned on $|P^{-1}(1)| = t$ (where $t \in \{1 \dots 2^k\}$ is some parameter), then P will satisfy the first condition (supporting a pairwise independent distribution) with high probability as long as $t > ck^2$ for some constant c [14] while it will *not* satisfy the second condition even for t as large as $\exp(k/5)$.

Another line of work has been concerned with proving *unconditional* lower bounds for these problems on restricted families of algorithms. These works considered convex relaxations for CSPs, where we say that a CSP is *approximation resistant* for some relaxation $\mathcal{R}$ if there is an instance for which a random assignment is essentially optimal, but the relaxation value is $1 - o(1)$ (namely, the relaxation "thinks" that it's possible to satisfy almost all constraints). Interestingly, the unconditional results match the conditional ones. That is, for certain weaker relaxations (namely, the Sherali-Adams linear programming hierarchy or Sherali-Adams augmented with the basic semidefinite program), there are unconditional results for the same predicates that were shown approximation-resistant under the UGC [30, 105, 89]. (This is of course not a coincidence, as the UGC is intimately connected with some of these weaker relaxations [92].) In contrast, for the stronger *Sum of Squares (SOS)* (also known as *Lasserre*) relaxation [100, 88, 90, 80], the previously known results [58, 95, 103] utilized the same conditions as in Chan's NP-hardness result (and in fact inspired Chan's work). This leads us to the following question of our interest:

Question 2. *Is there a pairwise independent predicate P such that $2^{o(n)}$ time Sum-of-Squares algorithm beat the random assignment for CSP(P)?*

In Chapter 3 of this thesis, we show that the pairwise independence condition suffices for lower bounds even for this stronger Sum-of-Squares hierarchy. This result is interesting in its own right and, based on past experience, could also be viewed as suggesting that it may be possible to improve the UGC-based results to results based on $\mathbf{P} \neq \mathbf{NP}$. Roughly speaking, we show that for every k and an arbitrarily small $\varepsilon > 0$, there

exists a set $\mathcal{I} = \{C_1, \ldots, C_m\}$ of k-tuples of literals (i.e. variables or their negations) over the variables $x_1, \ldots, x_n$ such that **(1)** for every assignment x to the variables, the induced distribution on $\{0,1\}^k$ obtained by taking a random $i \in [m]$ and looking at the literals in C_i is ε-close to the uniform distribution on $\{0,1\}^k$ but **(2)** for every pairwise independent distribution μ over $\{0,1\}^k$, there is a relaxation-solution that "cheats" the $\Omega(n)$-degree SOS relaxation to think that there is a distribution $\mathcal{D}$ over assignments (i.e. $\{0,1\}^k$) such that for every $i \in [m]$, the projection of $\mathcal{D}$ to the literals in C_i is distributed according to μ. This immediately implies that predicates supporting a pairwise independent distribution are approximation-resistant for this relaxation. We now formally state our results:

In Chapter 3, we show the following result that shows that predicates supporting pairwise independent distributions are approximation-resistant for $\Omega(n)$-degree SOS. This is based on joint work with Boaz Barak and Siu On Chan [20].

Theorem 1.1.4 (Lower Bound for Pairwise Independent CSPs)**.** *For every $\varepsilon > 0$ and $P : \{\pm 1\}^k \to \{0,1\}$, if there exists a pairwise independent distribution μ supported on $P^{-1}(1)$ then there exists $\delta > 0$ such that for all n there is a set $\mathcal{I} = \{C_1, \ldots, C_m\}$ of k-tuples of literals over $x_1, \ldots, x_n$ such that*

1. *For every $x \in \{\pm 1\}^n$, $\mathbb{E}_{C \in \mathcal{I}} P(C(x)) \leqslant \frac{|P^{-1}(1)|}{2^k} + \varepsilon$.*
2. *The value of the δn-degree Max-P SOS relaxation for the fraction of satisfiable constraints on the instance $\mathcal{I}$ is 1.*

Remark 1.1.5. The instance $\mathcal{I} = (C_1, \ldots, C_m)$ is actually obtained at random (with some pruning of a small fraction of the constraints, or alternatively, with some loss in the

"perfect completeness" condition). Thus the results above can also be thought as giving some evidence to a conjecture of Barak, Kindler and Steurer [25] that no polynomial-time algorithm (including in particular the SOS algorithm) can beat the basic semidefinite program on approximating random CSP instances.

1.2 Average Case Complexity and Planted Clique

Next, we shift attention to average case complexity and obtaining evidence of hardness for a central problem in average case complexity: *planted clique*.

Arising from the 1976 work of Karp [75], the problem was formally defined by Jerrum [73] and Kucera [79] as follows: given a random Erdős-Rényi graph G from the distribution $G(n, 1/2)$ where every edge is chosen to be included with probability $1/2$ independently of all others in which we *plant* an additional clique (i.e., set of vertices that are all neighbors of one another) S of size ω, find S. It is not hard to see that the problem can be solved by brute force search which in this case takes quasipolynomial time whenever $\omega > c \log n$ for any constant $c > 2$. Despite considerable effort, the best polynomial-time algorithms only work when $\omega = \varepsilon\sqrt{n}$, for any constant $\varepsilon > 0$ [7].

Over the years planted clique and related problems have found applications to important questions in a variety of areas including community detection [61], finding signals in molecular biology [1], discovering motifs in biological networks [86, 70], computing Nash equilibrium [64, 13], property testing [5], sparse principal component analysis [32], compressed sensing [77], cryptography [74, 8] and even mathematical finance [3].

Thus, the question of whether the currently known algorithms can be improved

is of great interest. Unfortunately, it is unlikely that lower bounds for planted clique can be derived from conjectured complexity class separations such as $\mathbf{P} \neq \mathbf{NP}$, precisely because it is an average-case problem [48, 34]. Our best evidence for its difficulty comes from showing limitations on powerful *classes* of algorithms. In particular, since many of the algorithmic approaches for this and related problems involve spectral techniques and convex programs, limitations for these types of algorithm are of significant interest. One such negative result was shown by Feige and Krauthgamer [46] who proved that the $n^{O(d)}$-time *degree d Lovász-Schrijver semidefinite programming hierarchy* (LS_+ for short) can only recover the clique if its size is at least $\sqrt{n/2^d}$.[1]

However, recently it was shown that in several cases, the *Sum-of-Squares (SOS) hierarchy* [100, 90, 80] can be significantly more powerful than other algorithms such as LS_+ [19, 22, 24]. In particular, it was conceivable that the SOS hierarchy might be able to find planted cliques that are much smaller than $\sqrt{n}$ in polynomial time, or at least be able to beat brute force search.

The first SOS lower bound for planted clique was shown by Meka, Potechin and Wigderson [85] who proved that the degree d SOS hierarchy cannot recover a clique of size $\tilde{O}(n^{1/d})$. This bound was later improved on by Deshpande and Montanari [43] for the special case of $d = 4$ to $\tilde{O}(n^{1/3})$.

However, prior works still left open the tantalizing possibility of even degree 4 SoS algorithm detecting $n^{0.49}$ size planted cliques:

[1] As we discuss later, formally such results apply to the incomparable *refutation* problem, which is the task of certifying that there is no ω-sized clique in a random $G(n, 1/2)$ graph. However, our current knowledge is consistent with these variants having the same computational complexity.

Question 1. *Can degree 4 algorithm from the SoS hierarchy detect $n^{0.49}$ size cliques planted in random graph $G \sim G(n, 1/2)$?*

1.2.1 A Tight Lower Bound at Degree 4

In Chapter 4 of this thesis, we answer the question above in the negative and show that the first non trivial extension of the spectral algorithm, namely the SoS algorithm of degree 4, cannot find cliques of size $\approx \sqrt{n}$, a bound optimal within $\operatorname{poly}\log(n)$ factors. Our lower bound for degree 4 is obtained by a careful "correction" to the certificate used by [85] and [43] in their lower bounds. This is based on joint work with Samuel Hopkins and Aaron Potechin [66]

Theorem 1.2.1 (Degree 4 Lower Bound)**.** *The canonical degree 4 SoS relaxation of the planted clique problem ((4.0.1)) has an integrality gap of at least $\tilde{O}(\sqrt{n})$ with high probability.*[2].

A similar result was obtained in an independent work by Raghavendra and Schramm [93]. We also give a tight analysis of the certificate considered by [85] and [43] and show that it yields a lower bound of $n^{\frac{1}{\frac{d}{2}+1}}$ for degree d SoS relaxation.

Theorem 1.2.2 (Tight Analysis of MPW)**.** *For every $d = o(\sqrt{\log(n)})$, the canonical degree d SoS relaxation for the planted clique problem ((4.0.1)) has an integrality gap of at least $\tilde{O}(n^{\frac{1}{\frac{d}{2}+1}})$.*

Curiously, the certificate used in [85, 43] is sufficient to show an $\Omega(\sqrt{n/2^d})$ lower bound for the degree d $LS+$ hierarchy [46] (which is a weaker SDP that also runs in time $n^{O(d)}$). However, a generalization of an argument of Kelner (see Section 4.8) shows that

[2]Throughout this thesis, we use $\tilde{O}$ to hide polylogarithmic factors in n

this is *not* the case for the SoS hierarchy, and our analysis of this certificate is tight. Thus, we can conclude that to get stronger lower bounds for higher degree SOS it is necessary and sufficient to utilize more complicated constructions of certificates than those used for weaker hierarchies.

1.2.2 A Nearly Tight Lower Bound at All Degrees

The above results still left open the possibility that the constant degree (and hence polynomial time) SoS algorithm can significantly beat the $\sqrt{n}$ bound, perhaps even being able to find cliques of size n^{ε} for any fixed $\varepsilon > 0$.

In Chapter 5, we resolve this question and show a strong lower bound that precludes any such algorithm. This is based on joint work with Boaz Barak, Sam Hopkins, Jonathan Kelner, Ankur Moitra and Aaron Potechin [21].

Theorem 1.2.3 (Optimal Planted Clique Lower Bound)**.** *There is an absolute constant c so that for every $d = d(n)$ and large enough n, the SOS relaxation of the planted clique problem has integrality gap at least $n^{1/2-c(d/\log n)^{1/2}}$.*

The main contribution of this work is to replace the ad-hoc certificate constructions in the previous works with a general principled method, that we call *pseudo-calibration*. In addition to recovering the previous certificate constructions for problems such as CSPs [59, 96, 37], our construction makes a concrete connection to the corresponding *planted vs random distinguishing* problems and provides a better intuition behind the limitations for SOS algorithms by viewing them from a "computational Bayesian probability" lens that is of its own interest. Moreover, there is some hope that this view could be useful not

just for more negative results but for SOS *upper bounds* as well. In particular our proof elucidates some aspects of the way in which the SOS algorithm is more powerful than the LS_+ algorithm.

Chapter 2

Sub-Exponential Lower Bounds on Linear Programs for MAX-CSP

In this chapter, we show a sub-exponential lower bound on linear programs for constraint satisfaction problems. The results of this chapter were obtained in joint work with Raghu Meka and Prasad Raghavendra [78].

The core of our result above is a new structural result about *rectangles* that has various applications in communication complexity in the context of *lifting* query lower bounds to communication lower bounds. We begin by an overview of the results presented in this chapter.

2.1 CSPs, Linear programming relaxations, Sherali-Adams hierarchy

A MAX-CSP (hereon referred to only as CSP) is defined by a predicate $P : \{-1,1\}^k \to \{0,1\}$. An instance of the CSP, $\mathcal{I}$, defined by a collection of m k-tuples of literals $C_1, C_2, \ldots, C_m$ on n Boolean variables $x_1, x_2, \ldots, x_n$. The algorithmic problem is to find an assigment $x = (x_1, x_2, \ldots, x_n) \in \{-1,1\}^n$ to maximize the number of satisfied constraints:

$$\mathrm{opt}(\mathcal{I}) = \max_{\S \in \{-\infty,\infty\}^\setminus} \sum_{\rangle=\infty}^{\Updownarrow} \mathcal{P}(C_\rangle(\S)) \equiv \max_{\S \in \{-\infty,\infty\}^\setminus} \mathcal{I}(\S), \tag{2.1.1}$$

where we define $\mathcal{I}(\S) = \sum_{\rangle=\infty}^{\Updownarrow} \mathcal{P}(C_\rangle(\S))$.

For instance, MAX-CUT corresponds to the case where the predicate $P : \{-1,1\}^2 \to \{0,1\}$ is defined by $P(a,b) = (1-ab)/2$ with instances corresponding to graphs.

Here we consider a broad-class of linear programming relaxations for CSPs obtained by *linearizing* the objective function $\mathcal{I}(\S)$. Formally, given a predicate P, and an integer D, we want:

Definition 2.1.1 (Linearization)**.** 1. A vector $v_x \in \mathbb{R}^D$ for every $x \in \{-1,1\}^n$.

2. A vector $w_{\mathcal{I}} \in \mathbb{R}^D$ for every instance $\mathcal{I}$ of the CSP.

3. The vectors v_x and $v_{\mathcal{I}}$ satisfy the condition that for every assignment x and every instance $\mathcal{I}$, $\mathcal{I}(\S) = \langle \sqsupseteq_{\mathcal{I}}, \sqsubseteq_{\S} \rangle$.

Given a linearization as above, a polytope $\mathcal{P} \subseteq \mathbb{R}^{\mathcal{D}}$ such that $\{v_x : x \in \{-1,1\}^n\} \subseteq \mathcal{P}$ defines a relaxation for the CSP as follows. Given an instance, we look at the linear program

$$\mathrm{opt}_{\mathcal{P}}(\mathcal{I}) = \max_{\dagger \in \mathcal{P}} \langle \sqsupseteq_{\mathcal{I}}, \dagger \rangle.$$

Clearly, $\mathrm{opt}(\mathcal{I}) \leqslant \mathrm{opt}_{\mathcal{P}}(\mathcal{I})$. The complexity of the relaxation is defined as the number of facets (or inequalities) needed to describe the polytope $\mathcal{P}$.

(c,s)-approximation. We say that $\mathcal{P}$ achieves an α-approximation factor if for every valid instance $\mathcal{I}$, $\mathrm{opt}_{\mathcal{I}} \geqslant \alpha\, \mathrm{opt}_{\mathcal{P}}(\mathcal{I})$. More generally, we say that $\mathcal{P}$ achieves a (c,s)-*approximation* if for all valid instances $\mathcal{I}$ such that $\mathrm{opt}(\mathcal{I}) \leqslant \int$, $\mathrm{opt}_{\mathcal{P}}(\mathcal{I}) \leqslant \rfloor$.

The above framework introduced in the work of [38] generalizes the *extended formulation* framework of Yannakakis [107] and its adaptation to approximation algorithms studied in [35]. We prove that despite its apparent vast generality, when it comes to CSPs, general linear programs as above are only as powerful as those obtained from the Sherali-Adams hierarchy:

Theorem 2.1.2. *There exist constants $0 < h < H < 1$ such that the following holds. Consider a function $f : \mathbb{N} \to \mathbb{N}$. Suppose that the $f(n)$-round Sherali-Adams relaxation for a CSP cannot achieve a (c, s)-approximation. Then, no LP relaxation of size $n^{hf(n)}$ can achieve a (c, s)-approximation for the CSP on n^H variables.*

In conjunction with known lower bounds for Sherali-Adams relaxations, the above theorem yields the following corollary:

Corollary 2.1.3. *For every $\varepsilon > 0$, there exist constants $c_\varepsilon, c'_\varepsilon$ such that the following holds. No LP relaxation of size less than $2^{n^{c_\varepsilon}}$ can achieve a $(7/8 + \varepsilon)$-factor approximation for MAX-3SAT. Similarly, no LP relaxation of size less than $2^{n^{c'_\varepsilon}}$ can achieve a $(1/2 + \varepsilon)$-factor approximation for MAX-CUT.*

The above results for CSPs are established through a more general claim on *non-negative rank* of a class of matrices that we refer to as *lifted matrices*. We explain this connection and results next.

2.1.1 Lifting degree lower bounds to rank lower bounds

In the seminal work introducing extended formulations, Yannakakis showed that the extended formulation complexity of an optimization problem is precisely the non-

negative rank of the associated slack matrix. In [], this connection was subsequently extended to approximation by linear programs. All known lower bounds on the size of extended formulations rely on this connection as do we.

Definition 2.1.4 (Non-negative Rank)**.** Let M be a non-negative matrix. The non-negative rank of M, denoted by $\mathsf{rank}_+(M)$ is the least positive integer r such that there exist non-negative rank 1 matrices M_i for $1 \leqslant i \leqslant r$ satisfying $M = \sum_{i=1}^{r} M_i$.

Proving lower bounds on non-negative rank of specific matrices is often non-trivial; the first breakthrough was achieved by the work of [50]. Before the work of [50] there was no super-constant separation between the rank and non-negative rank of any explicit matrix.

We give a tight characterization of the non-negative rank of a broad-class of matrices—*lifted matrices*—that were studied before in communication complexity[1].

The lifted matrix of f, a function on $\{-1,1\}^n$, is a matrix with rows and columns indexed by $\{-1,1\}^{bn} \times \{-1,1\}^{bn}$. Important to this mapping of f into a matrix is the notion of a b-bit gadget which is just a function $g : \{-1,1\}^b \times \{-1,1\}^b \to \{-1,1\}$ that maps a pair of b bit strings into a single bit. As in the work on pattern matrices before, we will end up using the Boolean inner product function as our gadget g, i.e., $g(x,y) = \Pi_{i=1}^{b}(1+x_i)(1+y_i)$.

We are now ready to define f-lifted matrix with a b-bit gadget g.

Definition 2.1.5 (f-Lifted Matrix)**.** Let $f : \{-1,1\}^m \to \mathbb{R}_+$ be a non-negative function. For positive integer parameter *blocksize* b, we define a non-negative matrix $M_f \in \{-1,1\}^{bn} \times$

[1] In [99], these matrices have been referred to as pattern matrices

$\{-1,1\}^{bn}$ as follows. We view $x \in \{-1,1\}^{bm}$ as m blocks of size b bits each and write x^i for i^{th} block of bits. For $x, y \in \{0,1\}^{bn}$, we write $g^{\otimes n}(x,y)$ to be the n bit output produced by applying g to each of the n corresponding disjoint blocks of x and y of b bits each. Then, for any $x, y \in \{0,1\}^{bm}$ we define

$$M_f(x,y) = f(g^{\otimes n}(x,y)).$$

Our main result gives a lower bound on the non-negative rank of M_f constructed with the Boolean inner product gadget of b bits in terms of a corresponding measure of f - that we next define.

Definition 2.1.6 (Juntas and Non-negative Degree)**.** A function $h : \{-1,1\}^n \to \mathbb{R}$ is a *d-junta* if it only depends on at most d coordinates. A function $h : \{-1,1\}^n \to \mathbb{R}_+$ is a *conical d*-junta if it can be written as a non-negative linear combination of non-negative d juntas.

For any $f : \{-1,1\}^n \to \mathbb{R}_+$, the non-negative degree of f, written as $\mathsf{deg}_+(f)$, is the least positive integer d such that f can be written as a conical d-junta.

We show that for any non-negative function f that cannot even be approximated by low-degree conical juntas, the corresponding matrix M_f has a high non-negative rank. Since we work with an approximation for f that works up to additive errors, we normalize f so that $\mathbb{E}[f] = 1$.

Theorem 2.1.7 ($\mathsf{rank}_+(M_f)$ vs $\mathsf{deg}_+(f)$)**.** *There exist constants $0 < c, C$ such that the following holds. Let $f : \{-1,1\}^m \to \mathbb{R}^+$ such that $\mathbb{E}[f] = 1$. Then, for all $\eta \geqslant c/m$ and M_f defined with*

blocksize $b \geqslant C(\log_2 m)$,

$$\mathsf{rank}_+(M_f) \geqslant 2^{\Theta(\mathsf{deg}_+(f+\eta)b)}.$$

Note that $\mathsf{rank}_+(M_f) \leqslant 2^{\mathsf{deg}_+(f)\cdot b}$. Thus, the above theorem is tight up to constant factors (in the exponent) and working with $\mathsf{deg}_+(f+\eta)$.

2.1.2 Decomposing high-entropy distributions

The technical core of all our results is a decomposition result for high-entropy distributions that may be of independent interest. Let X, Y be two independent distributions over $(\{0,1\}^b)^n$ with very high min-entropy: $H_\infty(X), H_\infty(Y) \geqslant (nb) - C$. That is, X, Y have min-entropy *deficiency* at most C. Such distributions arise commonly in communication complexity—they correspond to large rectangles. We would like to study the product distribution $X \times Y$ on $(\{0,1\}^b)^n$ and decompose them into distributions with more structure.

For instance, one class of such high min-entropy distributions are those where some C/b blocks of X are assigned to a fixed string and the rest are all uniformly random. That is, for some set $I \subseteq [n]$ with $|I| \leqslant C/b$, X_I is a fixed string whereas $X_{[n]\setminus I}$ is uniformly random over $(\{0,1\}^b)^{[n]\setminus I}$. Similarly, Y could satisfy a similar property for a set $J \subseteq [n]$ with $|J| \leqslant C/b$. An especially nice scenario is when X, Y are even *aligned* in the sense that $I = J$. Such distributions are very nice for studying communication complexity of lifted functions as the distribution $Z = g^n(X, Y)$ is very simple: some coordinates of Z are fixed, while the rest are uniformly random. We show that as long as X, Y have high min-entropy,

we can approximately decompose the product distribution $X \times Y$ into distributions that are essentially as simple and aligned as in the above discussion.

To this end, we next introduce the notion of *blockwise-dense* distributions; they were first defined in the work of [56] and play a crucial role here.

Definition 2.1.8. A distribution X on $\{0,1\}^{bn}$ is *blockwise-dense* if for every $I \subseteq [n]$, $H_\infty(X_I) \geqslant 0.9b|I|$.

Definition 2.1.9. A distribution X on $\{0,1\}^{bn}$ is a d-CBD ("conjunctive blockwise-dense") distribution if for some set of blocks $I \subseteq [n]$, $|I| \leqslant d$, $H_\infty(X_I) = 0$ and for every $J \subseteq [n] \setminus I$, $H_\infty(X_J) \geqslant 0.9b|J|$. We refer to d as the *degree* of the CBD distribution. We say two d-CBD distributions X, Y on $\{0,1\}^{bn}$ are aligned if the *fixed* blocks I are the same in both.

Our core lemma states that any two independent high-entropy distributions X, Y over $(\{0,1\}^b)^n$ can be approximated by a convex combination of aligned CBD distributions. The error of the approximation will depend on the entropy deficiency of $X \times Y$ and the degree of the CBD distributions used in the approximation.

Theorem 2.1.10. *There exists a constant $c > 0$ such that the following holds. Let $b \geqslant c \log n$ and X, Y be two independent distributions over $\{0,1\}^{bn}$ with $H_\infty(X)+H_\infty(Y) \geqslant 2nb-C$. Then, for all $d \geqslant cC/b$, $X \times Y$ can be written as a convex combination of distributions $X_1 \times Y_1, \ldots, X_N \times Y_N, E$ over $\{0,1\}^{bn} \times \{0,1\}^{bn}$, i.e., $X \times Y = \sum_{i=1}^{N} \lambda_i X_i \times Y_i + \lambda_{err} E$, such that*

- $0 \leqslant \lambda_1, \ldots, \lambda_N, \lambda_{err} \leqslant 1$, $\sum_{i=1}^{N} \lambda_i + \lambda_{err} = 1$.
- $|\lambda_{err}| < 2^{-db/c}$.
- *For every $1 \leqslant i \leqslant N$, X_i, Y_i are aligned d-CBD distributions.*

2.2 Preliminaries

2.2.1 Basic Notation

1. $\mathbb{P}_d^n$ be the collection of all polynomials of degree d on n variables on $\{-1,1\}^n$.

2. $\mathbb{1}(E)$ in the mean 1 indicator for the event E. That is, $\mathbb{1}(E)$ is 0 when E doesn't happen and $1/\mathbb{P}[E]$ when E happens.

3. For any function f, $\mathbb{E}[f]$ denotes the expectation of f on the uniform distribution over its domain.

4. For matrices M, $\mathbb{E}[M]$ denotes the expectation of $M(x,y)$ under x, y being uniformly random indices for its rows and columns.

5. For any $x \in \{-1,1\}^n$ and $K \subseteq [n]$, we write x_K to denote the projection of x on to the coordinates in K.

6. A *Boolean conjunction* on $\{-1,1\}^n$ is defined by a subset $I \subseteq [n]$ of variables and γ, an assignment to the variables in I and is the mean 1 indicator $\mathbb{1}(x_I = \gamma)$. Observe that we choose a non-standard scaling that satisfies: $\mathbb{E}[C] = 1$ for any conjunction C on $\{-1,1\}^n$.

7. For any $S \subseteq [n]$, the parity function $\chi_S(x) = \Pi_{i \in S} x_i$ for any $x \in \{-1,1\}^n$. Any function $f : \{-1,1\}^n$ has a Fourier expansion: $f(x) = \sum_{S \subseteq [n]} \hat{f}(S)\chi_S(x)$, for reals $\hat{f}(S)$ known as its Fourier coefficients.

2.2.2 Probability

Definition 2.2.1 (Density)**.** A function $p : \{-1,1\}^m \to \mathbb{R}^+$ is said to be a density if $\mathbb{E}[p(x)] = 1$.

At the cost of slight abuse of notation, we will use entropy for what is usually known as min-entropy.

Definition 2.2.2 (Entropy)**.** For a density p on $\{-1,1\}^m$, we write

$$H_\infty(p) = \min_{x \in \{-1,1\}^m} \log\left(2^m / p(x)\right)$$

for the (min)-entropy of the distribution associated with p.

2.2.3 Gadgets

Key ingredient to the transformation between a n bit function f and the corresponding lifted matrix $M_f \in \{-1,1\}^{bn} \times \{-1,1\}^{bn}$ is a *gadget* that we briefly discussed in the introduction. We note down some key definitions here.

A gadget $g : \{-1,1\}^b \times \{-1,1\}^b \to \{-1,1\}$ is just a function that maps a pair of b bit strings into a single bit. We can use g to product a function that operates on n disjoint pairs of b bit strings and outputs an n bit string. In this context, we will think of any string in $\{-1,1\}^{bn}$ as an object of the form $\{-1,1\}^{b^{\otimes n}}$ and for any $x \in \{-1,1\}^{bn}$, write x^i for the i^{th} block of b bits. We can then define $g^{\otimes n} : \{-1,1\}^{b^{\otimes n}} \times \{-1,1\}^{b^{\otimes n}} \to \{-1,1\}^n$ such that for any x, y, the i^{th} bit of $g^{\otimes n}(x,y) = g(x^i, y^i)$.

We will work with the Boolean inner product gadget in this paper: $g(x,y) = \Pi_{i=1}^{b}(1+x_i)(1+y_i)$ and when not explicitly noted, g will always denote the Boolean inner product function on pairs of b bit strings.

2.2.4 Lifted Matrices

For any function $f : \{-1,1\}^n \to \mathbb{R}$, we can use a b bit gadget g to obtain a f-lifted matrix $M_f \in \{-1,1\}^{bn} \times on^{bn}$ as defined in the introduction: for any $x, y \in \{-1,1\}^{bn}$, $M_f(x,y) = \lambda \cdot f(g^{\otimes n}(x,y))$, where λ is chosen to ensure that $\mathbb{E}[M_f] = 1$.

2.2.5 Sherali Adams Linear Program

Our results relate arbitrary linear programming relaxations for CSPs to the Sherali Adams relaxation that we now discuss.

We begin with the definition of a degree d pseudo-expectation [2]

Definition 2.2.3 (Sherali-Adams Pseudoexpectation)**.** A degree d Sherali-Adams pseudo-expectation, $\tilde{\mathbb{E}}$, is a linear operator on $\mathbb{P}^n_d$ such that

1. For every non-negative $p \in \mathbb{P}^n_d$ that depends on only d variables, $\tilde{\mathbb{E}}[p] \geqslant 0$, and
2. $\tilde{\mathbb{E}}[\mathbb{1}] = 1$.

Since $\tilde{\mathbb{E}}$ is a linear, it is completely described by its multilinear pseudomoments: $\tilde{\mathbb{E}}[\chi_S(x)]$ for $S \subseteq [n]$, $|S| \leqslant d$.

It is not hard to show that $\tilde{\mathbb{E}}$ of Definition 2.2.3 is equivalently described by a collection of local probability distributions μ_S, one for every $|S| \leqslant d$, $S \subseteq [n]$ such that for every $p \in \mathbb{P}^n_d$ that depends only on variables x_i such that $i \in S$, $\tilde{\mathbb{E}}[p] = \mathbb{E}_{\mu_S}[p]$.

[2] While there's no confusion in the present context, the term *pseudoexpectation* is more often used in connection with the Sum of Squares SDP hierarchy where in addition to the constraints in Definition 2.2.3, it satisfies the positive semidefiniteness constraint.

The Sherali-Adams linear programming relaxation of degree d solves the following optimization problem, given an instance $\mathcal{I}$ of CSP:

$$\max_{\tilde{\mathbb{E}}} \tilde{\mathbb{E}}[\mathcal{I}(x)], \tag{2.2.1}$$

where we see $\mathcal{I}$ as a polynomial of degree k, k being the arity of the associated predicate P and $\tilde{\mathbb{E}}$ ranges over all degree d-Sherali-Adams pseudoexpectations. We define $\mathsf{SA}_d(\mathcal{I})$ as the value of the the optimization problem (2.2.1).

It is not hard to show that the above optimization problem can be solved using a linear program on $n^{O(d)}$ variables and constraints for any $d \geqslant k$. This linear program is the Sherali-Adams linear program of degree d.

Sherali-Adams LP and Non-negative Degree:. Linear programming duality can be used to show the following:

Fact 2.2.4 (Sherali-Adams value and Non-negative Degree)**.** *Let $\mathcal{I}$ be an instance of CSP(P) with* $\mathrm{opt}(\mathcal{I}) = \int$*. The degree d-Sherali-Adams linear program achieves a value c for $\mathcal{I}$ if and only* $\mathsf{deg}_+(c - \mathcal{I}) \leqslant d$.

2.3 rank_+ of Lifted Matrices and LP Lower Bounds for CSPs

In this section, we show that proving a lower bound on the size of any linear program for CSP(P) reduces to proving non-negative rank lower bound on a f-lifted matrix for an appropriate choice of f. Formally, the goal of this section is to show:

Lemma 2.3.1 (LP Lower Bounds from Lifted Matrices)**.** *Let $c > s > 0$. Let $\mathcal{I}^*$ be an instance of*

CSP on n variables such that $\text{opt}(\mathcal{I}) = s$ *and let* $f(x) = c - \mathcal{I}(x)$*. For all* b*, any linear programming relaxation of CSP(P) that achieves* (c, s)*-approximation has size at least* $R \geqslant \Theta(\mathsf{rank}_+(M_f))$.

The proof of this lemma is simple. We first use the random planting idea (first employed by [38] and then later by [81]) to obtain a *random planting matrix* M such that any linear programming relaxation for CSP(P) of size R that achieves a (c, s)-approximation satisfies $R \geqslant \mathsf{rank}_+(M)$. Then, we show that M_f is a sub-matrix of M and therefore $\mathsf{rank}_+(M) \geqslant \mathsf{rank}_+(M_f)$. This completes the proof.

We now describe the two claims above in more detail. In the following, fix an instance $\mathcal{I}^*$ such that $\text{opt}(\mathcal{I}^*) = s$, and let $f(x) = c - \mathcal{I}^*(x)$.

2.3.1 Random Planting Matrix

We begin by setting up some notation to talk about the random planting matrix that we will define. Let $m = 2^b n$ for some positive integer parameter b. We define the *selection* function - this function takes input an index for each of the n blocks and a m bit string and outputs the n bit string obtained by choosing the single bit pointed by the index from each block. Formally,

Definition 2.3.2. Let b, n be positive integers and $m = 2^b n$. The selection function $s : (\{-1,1\}^b)^n \times \{-1,1\}^m \to \{-1,1\}^n$ is defined so that for any $x \in \{-1,1\}^n$ and $\mathsf{ind} \in (\{-1,1\}^b)^n$ and $i \in [n]$,

$$s(\mathsf{ind}, x)_i = x_{2^b * (i-1) + j},$$

where j is an integer in $\{0, 1, \ldots, 2^b - 1\}$ with bit representation ind_i.

We can now formally describe our random planting matrix:

Definition 2.3.3 (Random Planting Matrix)**.** The random planting matrix M has rows indexed by $(\{-1,1\}^b)^n$ and columns indexed by $\{-1,1\}^m$. For any row index ind and column index x, $M(\mathsf{ind}, x) = f(s(\mathsf{ind}, x))$.

We now explain what the random planting matrix is. For any $S \subseteq [m]$, such that $|S| = n$, we can construct an instance $\mathcal{I}_S$ of CSP(P) on m variables by arbitrarily associating each of the n variables in $\mathcal{I}^*$ with a distinct variable in S. Thus, as a function $\mathcal{I}_S$ will depend only on variables with indices in S out of $[m]$.

Any row index ind describes a set $S \subseteq [m]$ of size n: the set is constructed by thinking of m as n blocks of 2^b variables each and choosing one variable from each block. Calling this set $S(\mathsf{ind})$, it is then easy to see that $M(\mathsf{ind}, x) = c - \mathcal{I}_{S_{\mathsf{ind}}}(x)$. In other words, each row of the matrix M corresponds to an instance $\mathcal{I}$ of CSP(P) obtained by planting $\mathcal{I}^*$ over a subset of n variables chosen from $[m]$.

2.3.2 Reduction to Lifted Matrices

The non-negative rank of the matrix M is a lower bound on the size of any linear program that achieves a (c, s)-approximation for CSP(P). This is a consequence of the following fact, which follows from an application of Farkas' lemma (first shown by [107], see also [91], [35]) that relates the non-negative rank of the certain matrices and the size of linear programs that achieve (c, s)-approximation for CSP(P).

Fact 2.3.4 (Non-negative Rank and LP Size)**.** *Let $\mathcal{Q}$ be the collection of CSP(P) instances $\mathcal{I}$ such that* $\mathrm{opt}(\mathcal{I}) = s > 0$. *For $c > s$, let $\mathcal{M}_{=s}$ be the matrix with rows indexed by elements of $\mathcal{Q}$ and columns by $\{-1,1\}^n$ such that $M_{=s}(\mathcal{I}, x) = c - \mathcal{I}(x)$. Then, there exists an LP relaxation of*

CSP(P) of size R achieving a (c, s) approximation if and only if $\mathsf{rank}_+(M_{=s}) \leqslant R$.

Combining the above fact with the observation that for every $\mathcal{I}_S$ obtained by planting $\mathcal{I}^*$ on the subset of variable $S \subseteq [m]$, $|S| = n$, the $\mathrm{opt}(\mathcal{I}_S) = \mathrm{opt}(\mathcal{I}^*) = s$ and thus M is a sub-matrix of $M_{=s}$.

Corollary 2.3.5. *Suppose there exists an LP relaxation for CSP(P) of size R that achieves (c, s)-approximation. Then, $R \geqslant \Theta(\mathsf{rank}_+(M))$.*

The key technical lemma that reduces our job to lower bounding the non-negative rank of a lifted matrix is the following:

Lemma 2.3.6. *Let M be the random planting matrix constructed with positive integer parameters b, n, c from the instance $\mathcal{I}^*$ of CSP(P). Then,* $\mathsf{rank}_+(M) \geqslant \mathsf{rank}_+(M_f)$.

Proof. We will show that M_f is a submatrix of M. Specifically, let $q \in \{-1, 1\}^{bn}$. We think of q as n blocks of b bits each and write $q^1, q^2, \ldots, q^n$ to denote these b blocks. For each choice of $q \in \{-1, 1\}^{bn}$, we will associate a x_q such that $M_f(\mathsf{ind}, q) = M(\mathsf{ind}, x_q)$, yielding our claim.

We think of each $x \in \{-1, 1\}^{2^b n}$ as n blocks of 2^b bits each. Fix an arbitrary q and let x_q be the unique $x \in \{-1, 1\}^{2^b n}$ that satisfies for each i, $\langle \mathsf{ind}_i, q \rangle = x^i$ where the inner product on the LHS here is the standard bilinear form on GF(2). Observe that such an x exists and is unique to complete the proof. □

2.4 Juntas, Rectangles and Non-negative Rank of Lifted Matrices

In this section, we reduce the task of proving a lower bound on the non-negative rank of a f (i.e. Theorem 2.1.7) to showing the junta-approximation theorem which is the main technical contibution of the paper and the focus of the next section.

Important Parameters.

- $f : \{-1,1\}^n \to [0,1]$, and $\eta > 0$ such that $\mathsf{deg}_+(f+\eta) > D$.
- $b = \Theta(\log(n))$: blocksize.
- M_f: Lifted Matrix of f with blocksize b in $\{-1,1\}^m \times \{-1,1\}^m$ for $m = bn$.
- R: non-negative rank of the lifted matrix M_f.

We begin by restating the goal:

Theorem 2.4.1 ($\mathsf{rank}_+(M_f)$ vs $\mathsf{deg}_+(f)$, Theorem 2.1.7 restated)**.** *There exist constants $0 < c, C$ such that the following holds. Let $f : \{-1,1\}^n \to \mathbb{R}^+$ such that $\mathbb{E}[f] = 1$. Then, for all $\eta \geqslant c/n$ and M_f defined with blocksize $b \geqslant C \log_2 n$,*

$$\mathsf{rank}_+(M_f) \geqslant 2^{\Theta(\mathsf{deg}_+(f+\eta)b)}.$$

For matrix M_f, we define the function $A : \{-1,1\}^n \to \mathbb{R}^+$ by

$$A(z) = \mathop{\mathbb{E}}_{(x,y)\in g^{\otimes n^{-1}}(z)}[M_f(x,y)].$$

Observe for any $(x, y) \in g^{\otimes n^{-1}}(z)$, $M_f(x, y) = f(z)$. Then,

$$A(z) = f(z) \tag{2.4.1}$$

for every $z \in \{-1, 1\}^n$.

From here on, our proof can be broken into two broad steps. In the first step, our plan is to use the fact that A is obtained by averaging over $g^{\otimes n^{-1}}(z)$ combined with a bound $\mathsf{rank}_+(M_f)$ and obtain that A must be "close" to having non-negative degree at most d.

If we could ensure the stronger conclusion that A has small non-negative degree from above, then we would be immediately done. However, such a conclusion cannot hold and to salvage the situation, we need to work with a notion of closeness that goes well with the kind of approximation we obtain above. It turns out that this notion of closeness, while being far from the standard norms, still has a nice interpretation and corresponds to the correlation with an appropriately chosen function $\mathcal{L}$. To complete the proof, then, we will show that $\mathsf{deg}_+(f + \eta) > d$ yields an upper bound on $\mathcal{L}(f)$ while our approximation for A above yields a lower bound for $\mathcal{L}(A)$ taken together yielding a contradiction when R is too small.

Next, we begin with the details of the two steps above. Bulk of the technical work goes into accomplishing the first step above leading to bounds on error of approximation of certain functions associated with non-negative rank 1 matrices by non-negative juntas. We only state the result in the next subsect - the proof of this theorem is the central technical contribution of this paper and is the focus of the next section.

2.4.1 Junta Approximation Theorem

Consider the rank 1 matrix $pq^\intercal$ where p, q are non-negative vectors indexed by $\{-1,1\}^m$, satisfying $\mathbb{E}[p] = \mathbb{E}[q] = 1$. For any z, let

$$g^{\otimes n-1}(z) = \{(x,y) \in \{-1,1\}^m \times \{-1,1\}^m \mid g^{\otimes n}(x,y) = z\}$$

and let $\mathbb{1}_z : \{-1,1\}^m \times \{-1,1\}^m$ be the mean 1 indicator of $g^{\otimes n}(x,y) = z$:

$$\mathbb{1}_z(x,y) = \begin{cases} 2^n & \text{if } g^{\otimes n}(x,y) = z \\ 0 & \text{otherwise.} \end{cases} \tag{2.4.2}$$

Let $A_{p,q} = \mathbb{E}[\mathbb{1}_z(x,y)p(x)q(y)]$ be the appropriately scaled probability of (x,y)

We will need the notion of ε-decaying functions.

Definition 2.4.2 (ε-decaying functions)**.** A function $h : \{-1,1\}^n \to \mathbb{R}$ is said to ε-decaying if $\mathbb{E}[h] = 0$ and for every $I \subseteq [m]$, $|\hat{h}(I)| \leqslant \varepsilon^{|I|}$.

Next, for any $z \in \{-1,1\}^n$, we define $A_{p,q}(z)$ which is the probability under the product distribution defined by $p \times q$ on $\{-1,1\}^{bn} \times \{-1,1\}^{bn}$ of (x,y) satisfying $g^{\otimes n}(x,y) = z$.

Definition 2.4.3. Let Q be a non-negative matrix with rows and columsn indexed by $\{-1,1\}^{bn}$ such that $\mathbb{E}[Q] = 1$. Let $A_Q : \{-1,1\}^n \to [0,1]$ be defined by:

$$A_Q(z) = \mathbb{E}[\mathbb{1}_z(x,y)p(x)q(y)],$$

where $\mathbb{1}_z$ is the mean 1 indicator of $\{(x,y) \mid g^{\otimes n}(x,y) = z\}$ defined in (2.4.2). When $Q = pq^\intercal$ for densities $p, q \in \{-1,1\}^{bn}$, we write $A_{p,q}$ for A_Q.

Our junta approximation theorem says that for any densities p, q that have min-entropy at least $bn - t$, $A_{p,q}$ can be expressed as a perturbation of a non-negative junta of degree $d \approx t/b$.

Lemma 2.4.4 (Junta Approximation)**.** *Let p, q be densities over $\{-1,1\}^{bn}$ such that $H_\infty(p) + H_\infty(q) \geqslant 2bn - t$. Then, there exists a constant $\alpha_1 > 0$ such that for $d = \alpha_1 t/b$ and $\varepsilon = 2^{-0.5b}$, d-conjunctions $C_1, C_2, \ldots, C_N$, non-negative weights $\lambda_1, \ldots, \lambda_N, \lambda_{err} \geqslant 0$ such that $\sum_{i=1}^N \lambda_i + \lambda_{err} = 1$, ε-decaying functions $h_1, \ldots, h_N$ and a density γ on $\{-1,1\}^n$ such that*

$$A_{p,q}(z) = \sum_{i=1}^{N} \lambda_i C_i(z)(1 + h_i(z)) + \lambda_{err}\gamma,$$

for $\lambda_{err} \leqslant \varepsilon^d$.

2.4.2 A is an Approximate Conical Junta

The goal of this section is to use Lemma 2.4.4 and arrive at an approximation for A as a $\log(R)$-degree conical junta.

Lemma 2.4.5 (Conical Junta Approximation for A)**.** *Let $R = \mathsf{rank}_+(M_f)$. Then, there exist $J, \delta : \{-1,1\}^m \to \mathbb{R}^+$ satisfying*

$$A(z) = J(z) + \delta(z)$$

such that $d = \alpha_1 \log_2(R)/b$ and $\varepsilon = 2^{-\alpha_2 b}$ for some constants α_1, α_2:

1. *there exist conical d-juntas $C_1, C_2, \ldots, C_N$, ε-decaying functions h_i and non-negative constants λ_i such that $\sum_{i=1}^R \lambda_i = 1$ and*

$$J(z) = \sum_i \lambda_i c_i(1 + h_i),$$

2.

$$\mathbb{E}[\delta] \leqslant 2/R.$$

Proof. Since $\mathsf{rank}_+(M_f) = R$, there exists a collection of densities on $\{-1,1\}^{bn}$, $\{u_i \mid 1 \leqslant i \leqslant R\}$ and $\{v_i \mid 1 \leqslant i \leqslant R\}$ and a set of non-negative constants $\lambda_1, \lambda_2, \ldots, \lambda_R$ such that

$$M_f = \sum_{i=1}^{R} \lambda_i u_i v_i^\dagger.$$

Observe that

$$\sum_{i=1}^{R} \lambda_i = \sum_{i=1}^{R} \lambda_i = \sum_{i=1}^{R} \lambda_i \, \mathbb{E}[u_i v_i^\dagger] = \mathbb{E}[M_f] = 1.$$

Fix $t = 4\log_2(R)$. Let $Q \subseteq [R] = \{i \mid \|u_i\|_\infty \leqslant 2^{t/2} \text{ and } \|v_i\|_\infty \leqslant 2^{t/2}\}$. Then, $H_\infty(u_i) + H_\infty(v_i) \geqslant 2n - t$ for each $i \in Q$. Observe that for any $i \notin Q$, $\lambda_i \leqslant 2^{-t/2}$. This is because, $\lambda_i \, \mathbb{E}_y[u_i(x)v_i(y)] = \lambda_i u_i(x) \leqslant \mathbb{E}_y[M_f(x,y)] = \mathbb{E}[f] = 1$ and similarly, $\lambda_i v_i(y) \leqslant 1$ for any y.

Let

$$\delta_1(z) = \sum_{i \notin Q} \lambda_i \, \mathbb{E}[\mathbb{1}_z(x,y)u_i(x)v_i(y))])$$

and

$$J(z) = \sum_{i \in Q} \lambda_i \, \mathbb{E}[\mathbb{1}_z(x,y)u_i(x)v_i(y)].$$

In this notation, thus,

$$f(z) = \mathbb{E}[\mathbb{1}_z(x,y)M_f(x,y)] = J(z) + \delta_1(z).$$

From above, we know that for each $i \in Q$, $H_\infty(u_i) + H_\infty(v_i) \geqslant 2n - t$.

Applying Lemma 2.4.4 to u_i, v_i for every $i \in Q$, we have:

$$A_{u_i,v_i}(z) = \sum_{j=1}^{N} \lambda_{ij} C_{ij}(z)(1 + h_{ij}(z)) + \lambda_{i,err}\gamma_i, \tag{2.4.3}$$

where, $\lambda_{i,j}, 2^{-t} \geqslant \lambda_{i,err} \geqslant 0$, $\sum_j \lambda_{i,j} + \lambda_{i,err} = 1$, C_{ij}s are d-conjunctions, γ_i are densities and $h_{i,j}$ are ε-decaying for $\varepsilon = 2^{-0.5b}$.

Thus, $J(z) = \sum_{i\in Q,j} \lambda_{ij} \cdot \lambda_{ij} C_{ij}(1 + h_i)) + \sum_i \lambda_{i,err}\gamma_i$.

Define $\delta(z) = \sum_i \lambda_{i,err}\gamma_i(z) + \delta_1(z)$. Then, $\mathbb{E}[\delta] = \sum_{i,err} \lambda_{i,err} + \mathbb{E}[\delta_1]$

Now, $\mathbb{E}_z \mathbb{E}[\mathbb{1}_z(x, y)u_i(x)v_i(y)] = \mathbb{E}_{x,y}[u_i(x)v_i(y)] = 1$. Thus, $\mathbb{E}[\delta_1] = \sum_{i\notin Q} \lambda_i \leqslant 2^{-t/2}R \leqslant 1/R$. Similarly, $\sum_i \lambda_{i,err} \leqslant R2^{-t} \leqslant 1/R$. Thus, $\mathbb{E}[\delta] \leqslant 2/R$. □

2.4.3 Separating Function for Low Non-negative Degree Functions

Notation.

- $C_{\leqslant D}$: cone of non-negative D-juntas on $\{-1, 1\}^n$.
- $\mathcal{L} : \{-1, 1\}^n \to \mathbb{R}$: separating function.

In the previous section, we showed that there's a certain conical junta approximation for A. In this section, we will develop tools required to show that f cannot have an approximator of the form developed in the previous section if $\mathsf{deg}_+(f + \eta) > D = 2d$. That is, the main goal of this section is the following lemma:

Lemma 2.4.6. *Let d, D be positive integer parameters satisfying $d < 4D$ and for $\eta = 1/n$, let $\mathsf{deg}_+(f + \eta) > D$. Then, there is a degree D function $\mathcal{L}$ such that for any non-negative d-junta c satisfying $c(x) \leqslant 1$ for every x and any ε-decaying function h, for $\varepsilon < \frac{1}{n^4}$, $\mathbb{E}[\text{Ł}c(1+h)] \geqslant -\varepsilon^{3d} n^{4d}$, while $\mathbb{E}[\mathcal{L}(f + \eta)] < 0$.*

Lemma 2.4.7. *Suppose $D < \deg_+(f + \eta)$. There exists a degree D function $Ł : \{-1, 1\}^n \to \mathbb{R}$ such that:*

1. $\mathbb{E}[Ł] = \mathbb{E}[Ł \cdot 1] = 1$.
2. $\mathbb{E}[Ł f] < -\eta$.
3. $\mathbb{E}[Ł q] \geqslant 0$ *for every conical D-junta q on* $\{-1, 1\}^n$.
4. $|\hat{Ł}(S)| \leqslant 1$ *for every* $|S| \leqslant D$.

Proof. Observe that the set of conical $\leqslant D$-juntas denoted by $C_{\leqslant D}$ is convex. On the other hand from the hypothesis, we have that $f \notin C$. Thus, there exists a function $Ł : \{-1, 1\}^n \to \mathbb{R}$ such that $\langle Ł, h\rangle \geqslant 0$ for every $h \in C_{\leqslant d}$ but $\langle Ł, (f + \eta)\rangle < 0$. Moreover, since $C_{\leqslant D}$ is contained in the linear subspace of degree D polynomials, without loss of generality, we can assume that $Ł$ is also a degree D polynomial.

The first two properties are simple to verify. Since the constant function $\mathbb{1} \in C_{\leqslant D}$, we can assume (by rescaling, if needed) that $\langle Ł, 1\rangle = 1$ giving us the first property. Further, since $\langle Ł, f\rangle = Ł \cdot (f + \eta) - \langle Ł, \eta\rangle \leqslant -\eta$ giving us the second property.

For every subset $J \subset [m]$ of size $\leqslant D$ and a partial assignment $\alpha \in \{-1, 1\}^J$, the indicator function $\mathbb{1}[x_J = \alpha]$ is a non-negative junta and therefore satisfies $\langle Ł, \mathbb{1}[x_J = \alpha]\rangle \geqslant 0$ by our choice of $Ł$. If q is any non-negative D junta depending on variables with the indices contained in J, then, q is a non-negative linear combination of $\mathbb{1}[x_J = \alpha]$ and thus by linearity of inner product $\langle Ł, q\rangle \geqslant 0$ giving us the third property.

To obtain the bound on the Fourier coefficients of Ł, we have:

$$\left|\hat{L}(S)\right| = |\langle Ł, \chi_S(x)\rangle| = \left|\sum_{\alpha\in\{-1,1\}^S} \langle Ł, \chi_S(\alpha)\cdot \mathbb{1}[x_S = \alpha]\rangle\right| \leqslant \sum_{\alpha\in\{-1,1\}^S} \langle Ł, \mathbb{1}[x_S = \alpha]\rangle = 1$$

The inequality above uses the fact that for any $\alpha \in \{-1,1\}^S$, from part three proven above, $\langle Ł, \mathbb{1}[x_S = \alpha]\rangle \geqslant 0$. The last equality follows from the observation that for any $S \subseteq [bn]$, $\langle Ł, \sum_{\alpha\in\{-1,1\}^S} \mathbb{1}[x_S = \alpha]\rangle = \langle Ł, \mathbb{1}\rangle = 1$. □

The following technical property of Ł constructed in Lemma 2.4.7 will be required in our proof.

Lemma 2.4.8. *Let Ł be the separating function of degree D given by Lemma 2.4.7. Let c be a non-negative junta that depends only on variables $T \subseteq [n]$ and let S be any subset of $[n]$ such that and $|S| + |T| \leqslant D$. Then,*

$$\mathbb{E}[Łc\chi_S] \leqslant \mathbb{E}[Łc].$$

Proof. For $\mathbb{1}_{\chi_S(x)=1}$ and $\mathbb{1}_{\chi_S(x)=-1}$, the 0-1 indicators of the sets $\{x \mid \chi_S(x) = 1\}$ and $\{x \mid \chi_S(x) = -1\}$ respectively, we have:

$$\mathbb{E}[Łc\chi_S] = \mathbb{E}[Łc\mathbb{1}_{\chi_S(x)=1}] - \mathbb{E}[Łc\mathbb{1}_{\chi_S(x)=-1}] \tag{2.4.4}$$

Now, $c \cdot \mathbb{1}_{\chi_S(x)=-1}$ is a conical D-junta from the assumption in the statement. Thus, $\mathbb{E}[Łc\mathbb{1}_{\chi_S(x)=-1}] \geqslant 0$. Thus, (2.4.4) yields:

$$\mathbb{E}[Łc\chi_S] \leqslant \mathbb{E}[Łc\mathbb{1}_{\chi_S(x)=1}] + \mathbb{E}[Łc\mathbb{1}_{\chi_S(x)=-1}] = \mathbb{E}[Łc].$$

□

We are now ready to prove Lemma 2.4.6 which can be seen as a robust version of the property that $\mathbb{E}[\mathcal{L}c] \geqslant 0$ for every non-negative D-junta c.

Proof of Lemma 2.4.6. We use the separating function $\mathcal{L}$ constructed in Lemma 2.4.7. The second property is then a direct consequence of Lemma 2.4.7. For the first, we write $h = h_{low} + h_{high}$ where $h_{low} = \sum_{|S| \leqslant D-d} \hat{h}(S)\chi_S$, and $h_{high} = \sum_{|S|>D-d} \hat{h}(S)\chi_S$.

We have:

$$\mathbb{E}[Łc(1+h)] = \mathbb{E}[Łc(1+h_{low})] + \mathbb{E}[Łch_{high}].$$

Observe that since $|\hat{L}(S)| \leqslant 1$, $|Ł(x)| \leqslant n^D$. Noting that h is ε-decaying, c is bounded above by 1 and that h_{high} is a linear combination of parities of degree at least $D-d$, we have:

$$|\mathbb{E}[Łch_{high}]| \leqslant \varepsilon^{D-d} n^{2D}. \tag{2.4.5}$$

Next, noting that h_{low} is a linear combination of parities of degree at most d and that c is a function of degree at most d, we have, using Lemma 2.4.8, we have:

$$\begin{aligned} |\mathbb{E}[Łch_{low}]| &\leqslant \sum_{|S|\leqslant D-d} |\hat{h}(S)|\,|\mathbb{E}[Łc\chi_S]| \\ &\leqslant \sum_{|S|\leqslant D-d} |\hat{h}(S)|\,\mathbb{E}[Łc]. \end{aligned}$$

Since by definition $\mathbb{E}[h] = 0$, we have:

$$\mathbb{E}[Łc(1+h_{low})] \geqslant \mathbb{E}[Łc](1 - \sum_{1\leqslant|S|\leqslant D-d} \varepsilon^{|S|}) \geqslant \mathbb{E}[\mathcal{L}c](1-2Dn\varepsilon). \tag{2.4.6}$$

Using that $\varepsilon < 1/n^4$ and $D < n$, we have: $E[Łc(1+h_{low})] \geqslant 0$. Using (2.4.5) and (2.4.6),

$$\mathbb{E}[Łc(1+h)] \geqslant -\varepsilon^{3d} n^{4d}.$$

□

2.4.4 Putting Things Together

We can now complete the proof of Theorem 2.4.1.

Proof of Theorem 2.4.1. Recall: $R = \mathsf{rank}_+(M_f)$, A, the acceptance function of M_f and $\mathsf{deg}_+(f+\eta) > D$. Suppose $R > n^{8d}$. Using Lemma 2.4.5 we write: $A(z) = J(z) + \delta(z)$. Using Lemma 2.4.6, we have (using that $A = f$)

$$\begin{aligned}\mathbb{E}[Ł(f+\eta)] &= \mathbb{E}[\mathcal{L}A] + \eta \\ &= \mathbb{E}[ŁJ] + \mathbb{E}[Ł\delta] + \eta \\ &= \sum_i \lambda_i \, \mathbb{E}[Łc_i(1+h_i)] + \mathbb{E}[Ł\delta] + \eta.\end{aligned}$$

Using non-negativity of δ, we have:

$$\begin{aligned}&\geqslant (\sum_i \lambda_i)(-\varepsilon^{D-d} m^{2D}) - \|Ł\|_\infty \, \mathbb{E}[\delta] + \eta \\ &\geqslant -2^{-1.5bd} n^{4d} - 2n^{4d}/R + \eta.\end{aligned}$$

Choosing a large enough $b = \Theta(\log_2(n))$ for any $\eta > 1/n$, we thus have that $\mathbb{E}[\mathcal{L}(f+\eta)] > 0$.

On the other hand, from Lemma 2.4.7, we know that $\mathbb{E}[\mathcal{L}(f+\eta)] < -\eta + \eta = 0$ obtaining a contradiction. This completes the proof. □

2.5 The Junta Approximation Theorem

The goal of this section is to prove the junta approximation Lemma - Lemma 2.4.4:

Lemma 2.5.1 (Junta Approximation, Theorem 2.4.4 restated). *Let p, q be densities over $\{-1,1\}^{bn}$ such that $H_\infty(p) + H_\infty(q) \geqslant 2bn - t$. Then, there exist constants $\alpha_1, \alpha_2 > 0$ such that for $d = \alpha_1 t/b$ and $\varepsilon = 2^{-\alpha_2 b}$, d-conjunctions $C_1, C_2, \ldots, C_N$, non-negative weights $\lambda_1, \ldots, \lambda_N, \lambda_{err} \geqslant 0$ such that $\sum_{i=1}^N \lambda_i + \lambda_{err} = 1$, ε-decaying functions $h_1, \ldots, h_N$ and a density δ on $\{-1,1\}^n$ such that*

$$A_{p,q}(z) = \sum_{i=1}^{N} \lambda_i C_i(z)(1 + h_i(z)) + \lambda_{err}\delta,$$

for $\lambda_{err} \leqslant \varepsilon^d$.

2.5.1 Reduction to Decomposition of High Min-Entropy Distributions

Preliminaries and Notation.

- We use distributions, densities and random variables interchangeably with the meaning being clear from the context. We use entropy and min-entropy interchangeably.
- For a random variable Y on $\{-1,1\}^{bn}$, a block of variables $I \subseteq [n]$ and $\alpha \in \{-1,1\}^{bI}$, we write $(Y \mid Y_I = \alpha)$ for the random variable Y conditioned on the event $Y_I = \alpha$.
- For μ, a density on $\{-1,1\}^{bn} \times \{-1,1\}^{bn}$, and any subset $R \subseteq \{-1,1\}^{bn} \times \{-1,1\}^{bn}$, we write $\mu(R)$ for $\mathbb{P}_{(x,y)\sim\mu}[(x,y) \in R]$.
- For any sets $X, Y \subseteq \{-1,1\}^{bn}$, we write $\mu_{|X\times Y}$ for the distribution μ conditioned on $(x,y) \in X \times Y$.

Intuitively, one could imagine that high-entropy random variables on $\{-1,1\}^{bn}$ resemble the uniform random variable on $\{-1,1\}^{bn}$. The uniform density is constant and in particular, satisfies the claim of Lemma 2.5.1 with $d = 0$.

In order to build on this intuition and prove Lemma 2.5.1, we will decompose high-entropy densities p and q as a mixture of densities (i.e. a convex combination) each of which is a high-entropy density that satisfies some additional structure - we will call such densities *conjunctive blockwise dense* densities.

2.5.1.1 Conjunctive Blockwise Denseness and Junta Approximation

As before, we view $\{-1,1\}^{bn}$ as a collection of strings each of which consists of n disjoint blocks of b bits each. For a random variable X on $\{-1,1\}^m$ and any $I \subseteq [n]$, we write X_I for the random variable that ignores all but the blocks indexed by $i \in I$.

We begin with *blockwise dense* densities. These are random variables on $\{-1,1\}^{bn}$ such that the marginal on any sub-collection of blocks has high min-entropy. Intuitively, they are our versions of "almost uniform" densities.

Definition 2.5.2 (Blockwise Dense Random Variables)**.** A random variable X on $\{-1,1\}^{bn} = \{-1,1\}^{b^{\otimes n}}$ is said to be S-blockwise dense for $S \subseteq [n]$ if for every $I \subseteq S$, $H_\infty(X_I) \geqslant 0.9b|I|$. X is said to be blockwise dense if it is $[n]$-blockwise dense.

The following lemma (with a proof using Chor-Goldreich [40], similar to that of Lemma 13 in [56]) gives a concrete reason why blockwise dense random variables resemble the uniform distribution.

Lemma 2.5.3. *Suppose X and Y are independent, blockwise dense random variables in $\{-1,1\}^{bn}$ for $b > 7$. Let ν be the density of the random variable $g^{\otimes n}(X,Y)$ on $\{-1,1\}^n$. Then, for any $S \subseteq [n]$, $\nu = 1 + h$ for an an ε-decaying function h where $\varepsilon = 2^{-0.5b}$.*

We will need the following fact in the proof. Recall that g is the b bit Boolean inner product function with outputs in $\{-1, 1\}$.

Fact 2.5.4 (Chor-Goldreich [40])**.** *Suppose X, Y are independent, blockwise dense random random variables over $\{-1,1\}^{bn}$ for $b > 7$. Then, for any $I \subseteq [n]$, $|\mathbb{P}[\Pi_{i\in I} g(X_{\{i\}}, Y_{\{i\}}) = +1] - 1/2| \leqslant 2^{-0.8b|I|}$.*

Proof of Lemma 2.5.3. Observe that using Fact 2.5.4, we have $\hat{\nu}(S) = \mathbb{E}[\nu(z)\chi_S(z)] = \mathbb{E}_{z\sim\nu}[\chi_S(z)] = \mathbb{E}[\Pi_{i\in I} g(X_{\{i\}}, Y_{\{i\}})] \leqslant 2^{-0.8b|S|+2}$.

Let $h(z) = \sum_{|S|\geqslant 1} \hat{\nu}(S)\chi_S(z)$. Then, $\nu = 1 + h$ and by the above estimate, h is ε-decaying for $\varepsilon = 2^{-0.8b+2} \leqslant 2^{-0.5b}$. □

A conjunctive blockwise dense random variable is a random variable such that there is a small subset of blocks that is "fixed" (and thus has 0 entropy) and the the marginal on all other blocks is blockwise dense.

Definition 2.5.5 (d-CBD and aligned d-CBD pairs)**.** A random variable X on $\{-1,1\}^{bn}$ is *d-conjunctive blockwise dense* (*d-CBD*) if there exists a set of blocks $I \subseteq [n]$, $|I| \leqslant d$ such that $H_\infty(X_I) = 0$ and for every $J \subseteq [n] \setminus I$, $H_\infty(X_J) \geqslant 0.9b|J|$. We call such an I above the *fixed blocks* of X.

We say that two random variables X and Y with d-CBD densities on $\{-1,1\}^{bn}$ are *aligned* if they have the same fixed blocks.

Lemma 2.5.3 showed that if X, Y are $blockwise - dense$ random variables, then he density of the $g^{\otimes n}(X, Y)$ is an ε-decaying perturbation of the uniform density. In the

following, we show a refinement of Lemma 2.5.3 when X, Y are aligned d-CBD random variables - specifically, that $g^{\otimes n}(X, Y)$ is a an ε-decaying perturbation of a non-negative d-junta.

Lemma 2.5.6. *Let X, Y be aligned d-CBD random variables over $\{-1, 1\}^{bn}$ for $b > 7$ with the aligned blocks $I \subseteq [n]$ and $X_I = \alpha, Y_I = \beta$. Let ν be the density of $z = g^{\otimes n}(X, Y)$. Then, for $\varepsilon = 2^{-0.5b}$, there exists an ε-decaying function h such that $\nu = 2^{|I|}\mathbb{1}(z_I = g^{\otimes |I|}(\alpha, \beta))(1 + h)$.*

Proof. Observe that $z_I = g^{\otimes |I|}(\alpha, \beta)$ and $z_{\bar{I}} = g^{\otimes |\bar{I}|}(X_{\bar{I}}, Y_{\bar{I}})$. Since $X_I = \alpha$ and $Y_I = \beta$ are fixed, we have that $Z_I = g^{\otimes |I|}(\alpha, \beta)$. In particular, the density of z can be written as $\nu_I \cdot \nu_{\bar{I}}$ where ν_I is the density of z_I and $\nu_{\bar{I}}$ the density of $z_{\bar{I}}$.

Now, by definition, $X_{\bar{I}}, Y_{\bar{I}}$ are d-CBD random variables and thus, the density $\nu_{\bar{I}}$ of $z_{\bar{I}}$ can be written as $1 + h$ for an ε-decaying function for $\varepsilon = 2^{-0.5b}$ using Lemma 2.5.3. This completes the proof. □

Observe that by definition, d-CBD distributions have min-entropy at least $0.9bn - bd$ which for $d \ll n$, we consider high. Thus, any convex combination of d-CBD random variables is also a random variable with high min-entropy.

One could ask for a converse at this point: can every high min-entropy random variable be written as a convex combination of CBD random variables? In the next section, as a warmup, we show that the answer is indeed yes.

2.5.1.2 Decomposition Theorem and Proof of Lemma 2.5.1

For our applications, however, we need a significantly strengthened statement that gains from having a product of two high min-entropy random variables. Our main decomposition lemma states that up to some small error any pair of high min-entropy distributions p, q on $\{-1,1\}^{bn}$ can be decomposed into a convex combination of *aligned d-CBD* distributions.

Lemma 2.5.7. *Let p_0, q_0 be densities on $\{-1,1\}^{bn}$ satisfying $H_\infty(p_0) + H_\infty(q_0) \geqslant 2nb - t$ for some $t > 0$. Then, for all $d \geqslant 40t/b$ and $b > 80\log_2(n)$, there exists densities v_i such that $p_0 q_0 = \sum_{i=1}^{N} \lambda_i v_i + \lambda_{err} E$ such that*

- *for every $1 \leqslant i \leqslant N$, v_i is an aligned d-CBD density.*
- *$0 \leqslant \lambda_i, \lambda_{err} \leqslant 1$, $\sum_{i=1}^{N} \lambda_i + \lambda_{err} = 1$,*
- *$|\lambda_{err}| < 2^{-t}$, and*

It is tempting to attempt to prove this lemma by individually decomposing each of the two high min-entropy densities and then just take a convex combination of the pieces. However, it is unclear how to simultaneously ensure the property of each of the pieces in the decomposition being *aligned* and at the same time, having the error decrease exponentially int the entropy gap t. Having both these properties together is crucial to our proof.

It is easy to complete the proof of Theorem 2.5.1 using Lemma 2.5.6 and Lemma 2.5.7.

Proof of Theorem 2.5.1. Recall that $A_{p,q}(z) = \mathbb{E}[\mathbb{1}_z(x,y)p(x)q(y)]$ where z is the 2^n scaled 0-1 indicator of $\{(x,y) \mid g^{\otimes n}(x,y) = z\}$. Apply Lemma 2.5.7 to p and q. Let $X_i, Y_i, \lambda_i, \nu_i$ for $1 \leqslant i \leqslant N$ and $\lambda_e rr, E$ be the output parameters of the decomposition of pq.

Then, $A_{p,q}(z) = \sum_{i=1}^{N} \lambda_i \mathbb{E}[\mathbb{1}_z(x,y)\nu_i(x,y)] + \lambda_{err} \mathbb{E}_{(x,y)\sim E}[\mathbb{1}_z(x,y)E(x,y)]$. Let δ be the density induced on $\{-1,1\}^n$ by E,i.e., $\delta(z) = \mathbb{E}[\mathbb{1}_z(x,y)E(x,y)]$.

Observe that $\mathbb{E}[\mathbb{1}_z(x,y)\nu_i(x,y)]$ is just the density of $g^{\otimes n}(X_i, Y_i)$ and since (X_i, Y_i) are aligned d-CBD, using Lemma 2.5.6 we obtain for each $1 \leqslant i \leqslant N$, non-negative d-conjunctions J_i, coefficients C_i and ε-decaying functions h_i for $\varepsilon = 2^{-0.5b}$ such that $\mathbb{E}[\mathbb{1}_z(x,y)\nu_i(x,y)] = C_i J_i(z)(1 + h_i(z))$.

This completes the proof. □

2.5.2 Warm Up: Decomposing a Single Density

As a warm up, we show a simple short argument that any high entropy density can be decomposed into a convex combination of densities each of which are CBD . In the next section, we prove Theorem 2.1.10 which requires a much more subtle argument.

Lemma 2.5.8. *Any distribution X on $\{-1,1\}^{bn}$ with $H_\infty(X) \geqslant nb - t$ can be written as a convex combination of distributions, $X = \sum_{i=1}^{N} \lambda_i X_i + \lambda_{err} X_{err}$, such that*

- $0 \leqslant \lambda_i, \lambda_{err} \leqslant 1$, $\sum_{i=1}^{N} \lambda_i + \lambda_{err} = 1$.
- $|\lambda_{err}| \leqslant 2^{-t}$.
- *Each X_i is a $20t/b$-CBD distribution.*

Proof. We give an algorithm that obtains the claimed decomposition.

Decompose(Y).
Input. Y, a random variable on $\{-1,1\}^{bn}$ such that $H_\infty(Y) \geqslant nb - t$. $t' = 2t$.

1. If Y is blockwise-dense Terminate and return Y.
2. If $(H_\infty(Y)) \leqslant nb - t'$, Terminate and return Y, labeled as Error.
3. If not, let $I \subseteq [n]$ be a *maximal* set such that $H_\infty(Y_I) < 0.9b|I|$. Let $\alpha \in \{0,1\}^{b|I|}$ be such that $p = \mathsf{Pr}[Y_I = \alpha] \geqslant 2^{-0.9b|I|}$. Let $Y_1 \leftarrow (Y|Y_I = \alpha)$ and $Y_2 \leftarrow (Y|Y_I \neq \alpha)$. Then, Y_1 is blockwise-dense.
4. Return $p \cdot Y_1 + (1-p)$Decompose(Y_2).

To get the desired decomposition of X, we call Decompose(X).

For the analysis, first observe that in step (3) of the algorithm the entropy gap $|I|0.1b \leqslant t'$ and thus $|I| \leqslant 10t'/b$. Therefore, all Y_i that are not labeled Error by the algorithm are d-CBD for $d = 10t'/b$.

Finally, note that if the procedure terminates by labeling some Y_i as Error, then, since $H_\infty(Y) \geqslant nb - t$ while $H_\infty(Y_i) \leqslant nb - 2t$, the coefficient of Y_i in the decomposition is at most $2^{t-t'} = 2^{-t}$. □

2.5.3 Proof of Lemma 2.5.7: Decomposing Product of Two Densities

Now, suppose we have two independent random variables X^0, Y^0 over $\{-1,1\}^{bn}$ with total min-entropy at least $2nb - t$. A natural idea would be to apply Lemma 2.5.8 to X^0, Y^0 individually to decompose them into CBD sources (sans some error); however, the resulting components while blockwise-dense may not be aligned which is critical for our applications.

To prove Theorem 2.1.10, we will analyze the following recursive decomposition algorithm.

Let μ be a product density on $\{-1,1\}^{bn} \times \{-1,1\}^{bn}$. Below, we give a recursive algorithm that decomposes μ as a weighted combination of CBD densities whenever μ has high enough entropy.

2.5.3.1 The Decomposition Algorithm

Overview. The algorithm recursively partitions the domain $\{-1,1\}^{bn} \times \{-1,1\}^{bn}$ into rectangles $R_1, R_2, \ldots$. Recall that a rectangle in $\{-1,1\}^{bn} \times \{-1,1\}^{bn}$ is any subset of the form $X \times Y$ for $X, Y \subseteq \{-1,1\}^{bn}$. Let $\mu_{|R}$ be the density μ conditioned on $(x, y) \in R$. Then, any partition $R_1 \cup R_2 \cup \cdots \cup R_p = \{-1,1\}^{bn} \times \{-1,1\}^{bn}$ corresponds to writing μ as a convex combination:

$$\mu = \sum_{i=1}^{p} \mu(R_i)\mu_{|R_i}.$$

The algorithm will thus succeed if, given that $H_\infty(\mu) \geqslant 2nb - t$, there is a $p' \leqslant p$ such that for every rectangle R_i, $i \leqslant p'$, $\mu_{|R_i}$ is an *aligned-d*-CBD density and the error $\sum_{i=p'+1}^{p} \mu(R_i) \leqslant 2^{-\Omega(t)}$.

The algorithm obtains a partition of $\{-1,1\}^{bn} \times \{-1,1\}^{bn}$ in a very natural way with an explicit label identifying if a rectangle belongs to the error region or not. For every rectangle R that is not labeled as belonging to the error region, the set up will ensure that $\mu_{|R}$ is *aligned* d-CBD. The tricky part of the analysis would be to show that the total mass of the error region is $2^{-\Omega(t)}$ as required.

For exposition, we divide the algorithm into three subroutines Decompose ,

XDecompose and YDecompose and construct a labeled execution tree for it to aid us in the analysis. Decompose is the main procedure and takes input a rectangle $X \times Y \subseteq \{-1,1\}^{bn} \times \{-1,1\}^{bn}$. This rectangle will always satisfy the invariant of having an *aligned* set of nodes *fixed* - i.e. there's an explicitly identified block $U \subseteq [n]$ such that x_U, y_U for any $(x, y) \in X \times Y$ are set to some fixed values α, β (possible unequal). Observe that in the beginning, $X = Y = \{-1,1\}^{bn}$ and $U = \emptyset$.

Each time Decompose is invoked by the algorithm, we create a new node in the execution tree and identify it as being created by Decompose by indexing it with $v_1, v_2, \ldots,$. If the set of fixed blocks U in the input rectangle R has size $\geqslant d$, the algorithm terminates and adds R to Error_d - Error_d contains the set of rectangles that account for error owing to the number of fixed blocks in them exceeding d. Next, if $\mu(R) < \delta$, then R is labeled Error - Error maintains the collection of rectangles that are labeled as error because they had too small *measure*.

From now, assume that the input rectangle R qualifies to be in neither of Error or Error_d. If $\mu_{|R}$ is $blockwise - dense$, then, we terminate the algorithm and return R. Otherwise, there must be a block S and some assignment to variables in I, say α_I such that $\mathbb{P}_{\mu_{|R}}[X_I = \alpha_I] > 2^{-0.9b|I|}$ (or $\mathbb{P}_{\mu_{|R}}[X_I = \alpha_I] > 2^{-0.9b|I|}$). In this case, we call the subroutine XDecompose (or YDecompose , respectively) and create a new node for this call in the execution tree. We identify nodes that are associated with calls of XDecompose or YDecompose by labeling them as $w_1, w_2, \ldots$. The algorithm then continues by setting R to be the rectangle $X_{|S \neq \alpha_S} \times Y$ (or $X \times Y_{|S \neq \alpha_S}$, respectively).

The subroutine XDecompose (the case of YDecompose is analogous) takes input the rectangle $X \times Y$ along with the fixed block U that was the current input of the Decompose

routine when XDecompose was invoked in addition to the "too probable" X-block S found. XDecompose then chooses every possible value β for Y_S and for each β, calls Decompose recursively with the rectangle $X \times Y_{|S=\beta}$ and the fixed blocks $U \cup S$.

Decomposition Algorithm

Decompose.
Input A density μ on $\{-1,1\}^{bn} \times \{-1,1\}^{bn}$. A rectangle $X \times Y \subseteq \{-1,1\}^{bn} \times \{-1,1\}^{bn}$, $U \subseteq [n]$: subset of "fixed" blocks.
Parameters: $\delta \in (0,1)$, $d \in \mathbb{N}$. **Invariant:** $H_\infty(X_U) = H_\infty(Y_U) = 0$.

1. Initialize $\mathsf{Error}_d, \mathsf{Error} \leftarrow \emptyset$.
2. If $|U| \geqslant d$, Terminate after setting $\textsc{Error}_d \leftarrow \textsc{Error}_d \cup X \times Y$.
3. Set $R \equiv R_0 \equiv X \times Y$.
4. While $\mu_{|R_0}(R) \geqslant \delta$ do
   (a) If $\mu_{|R}$ is $\overline{U}$-blockwise-dense Terminate.
   (b) Else, if there is an $S \subseteq [n] \backslash U$ and $\alpha \in \{0,1\}^S$ such that
   $$\mathop{\mathbb{P}}_{\mu_{|R}}[X_S = \alpha] > 2^{-0.9b|S|},$$
   call YDecompose on input $(X_{|S=\alpha}, Y, U, S)$ and set $R \leftarrow X_{|S\neq\alpha} \times Y$
   (c) Else, if $\mathbb{P}_{\mu_{|R}}[Y_S = \alpha] > 2^{-0.9b|S|}$ then call XDecompose on input $(X, Y_{|S=\alpha}, U, S)$ and set $R \to X \times Y_{|S\neq\alpha}$.
5. Set $\mathsf{Error} \leftarrow \mathsf{Error} \cup R$.

YDecompose.
Input:. A rectangle $X \times Y \subseteq \{-1,1\}^{bn} \times \{-1,1\}^{bn}$, $U \subseteq [n]$: common subset of "fixed" blocks.
S: the set of newly fixed blocks in X but not in Y.
Invariant: $H_\infty(X_{U\cup S}) = H_\infty(Y_U) = 0$.

1. For every $\beta \in \{0,1\}^S$ DECOMPOSE$(X, Y_{|S=\beta}, U \cup S)$.

XDecompose.
Input:. A rectangle $X \times Y \subseteq \{-1,1\}^{bn} \times \{-1,1\}^{bn}$, $U \subseteq [n]$: common subset of "fixed" blocks.
S: the set of newly fixed blocks in Y but not in X.

Invariant: $H_\infty(X_U) = H_\infty(Y_{U\cup S}) = 0$.

1. For every $\beta \in \{0,1\}^S$ DECOMPOSE$(X_{|S=\beta}, Y, U \cup S)$.

The following is the main claim that captures the relevant claims about the algorithm.

Lemma 2.5.9 (Analysis of Decomposition Algorithm)**.** *Let μ be a density on $\{-1,1\}^{bn} \times \{-1,1\}^{bn}$ such that $H_\infty(\mu) \geqslant 2nb - t$. Let $R_1, \ldots, R_q, \mathsf{Error}_d, \mathsf{Error}$ be the partition of $\{-1,1\}^{bn} \times \{-1,1\}^{bn}$ returned by Algorithm 2.5.3.1 running with parameters δ, d. Then,*

1. *For each $i \leqslant q$, $\mu_{|R}$ is d-CBD.*
2. *$\mu(\mathsf{Error}) \leqslant d\delta$, and*
3. *$\mu(\mathsf{Error}_d) \leqslant 2^{-0.1bd} 2^t \cdot \left(\lceil \log 1/\delta \rceil + 2\right)^d$.*

We can complete the proof of Lemma 2.5.7 using Lemma 2.5.9.

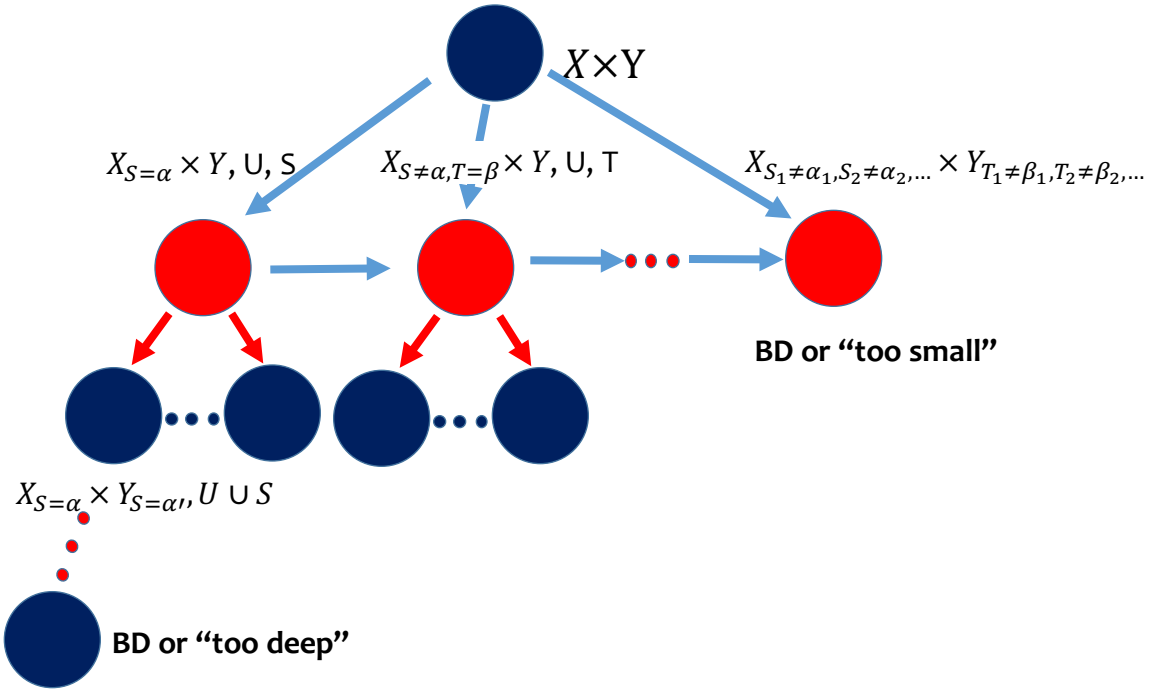


Figure 2.1: Execution Tree: Blue Nodes are Calls of Decompose , Red Nodes are calls of XDecompose or YDecompose . "too deep" ≡ Error, "too small" ≡ Error_d and BD = $blockwise - dense$

Proof of Lemma 2.5.7. We let $\mu = pq$. Then, $H_\infty(\mu) \geqslant 2nb - t$. We run Algorithm 2.5.3.1 on $\{-1,1\}^{bn} \times \{-1,1\}^{bn}$ with density μ and parameters $d = 40t/b$ and $\delta = 2^{-t}$.

We let E be the density $\mu_{|\text{Error}\cup\text{Error}_d}$ and for each i, let ν_i be the density $\mu_{|R_i}$ and let $\lambda_i = \mu(R_i) \geqslant 0$ and $\lambda_{err} = \mu(E) \geqslant 0$. Then, since $R_1, R_2, \ldots, R_q, \text{Error}, \text{Error}_d$ is a partition of $\{-1,1\}^{bn} \times \{-1,1\}^{bn}$ we have:

$$\mu = \sum_{i=1}^{q} \lambda_i \mu_{|R_i} + \lambda_{err} \mu_{|E},$$

and $\sum_{i=1}^{q} \lambda_i + \lambda_{err} = 1$.

Each ν_i is d-CBD from claim 1 of Lemma 2.5.9. The last two claims of Lemma 2.5.9 can now be used to bound λ_{err}:

$$\lambda_{err} \leqslant d2^{-t} + 2^{-0.1Ct}2^t 2^{Ct/b \log(t+2)}.$$

Now, $\log(t) \leqslant 2\log(n)$ for any $b < n$. Choose $C = 40$ and we know that $n > b > 80\log(n)$. Then, the second term above evaluates to at most 2^{-2t}. The first term is at most 2^{-t}. Thus the $\lambda_{err} < 2^{-t}$ as required.

□

2.5.3.2 Analysis:Proof of Lemma 2.5.9

In the following, we analyze the procedure above.

Execution Tree. Consider the execution tree of the decomposition algorithm (Figure ??). A node ρ corresponding to a call of Decompose is then associated with parameters X_ρ, Y_ρ and U_ρ for $X_\rho, Y_\rho \subseteq \{-1,1\}^{bn}$ and $U_\rho \subseteq [n]$. Nodes ρ corresponding to a call of XDecompose or YDecompose are associated with an additional parameter of $S_\rho \subseteq [n]$. The calls of Decompose and XDecompose or YDecompose alternate.

As stated in the overview, the tree yields a partition $R_1 \cup R_2 \cup \cdots R_q$ of $\{-1,1\}^{bn} \times \{-1,1\}^{bn}$ where each R_i are rectangles. The final decomposition can be obtained by writing

$$\mu = \sum_{i=1}^{q} \mu(R_i)\mu_{|R_i}.$$

Before we analyze the decomposition, we set some notation.

- For each vertex ρ, let $\mu(\rho) = \mu(X_\rho \times Y_\rho)$.
- The $\mathsf{depth}(\rho)$ of any node ρ in the tree is the number of edges on the path from root to ρ.
- For a child b of a vertex a, let $\mu(a \mid b) = \mu(b)/\mu(a)$ - the conditional probability of traversing the edge (a, b) in the tree when at vertex a.
- We will reserve the letter v (and various suffixes) for nodes corresponding to calls of Decompose (i.e., the vertices in odd-layers) and the letter w for nodes corresponding to calls of XDecompose or YDecompose (i.e., the vertices in even-layers).
- For a vertex v let $C_v = \{w_1, \ldots, w_{c_v}\}$ be the children of v, where we assume that $w_1, \ldots, w_{c_v-1}$ lead to recursive calls of XDecompose or YDecompose while w_{c_v} corresponds to the rectangle marked as Error in Step(4) of Decompose .
- We say a node v in the tree is *bad* if it is at depth $2d$ but $\mu_{|X_v \times Y_v}$ is not blockwise-dense in the non-fixed blocks. A root-to-leaf path $v_1, w_1, v_2, w_2, \ldots, v_d, w_d, v_{d+1}$ is called a *bad path* if v_{d+1} is bad.

We can now begin our analysis. It is immediate from the algorithm that all the leaves in the execution tree correspond to the calls of Decompose and further, if for such a leaf v, the corresponding rectangle $R(v)$ is not added to Error or Error_d at any point, then $\mu_{|R}$ is d-CBD. We record this observation as the following fact:

Lemma 2.5.10 (Good Rectangles)**.** *Let v be a leaf in the tree with the corresponding rectangle $R(v)$ such that $R(v) \notin \mathsf{Error} \cup \mathsf{Error}_d$. Then, $\mu_{|R}$ is d-CBD.*

Proof. The assumptions imply that $\mathsf{depth}(v) \leqslant 2d$ and thus the fixed blocks U in R satisfy $|U| \leqslant d$. Since $R(v)$ is not added to Error, $R(v)$ is a leaf because $\mu_{|R}$ is $\overline{U}$-blockwise dense or equivalently, $\mu_{|R}$ is d-CBD □

It is also easy to bound $\mu(\mathsf{Error})$:

Lemma 2.5.11. $\mu(\mathsf{Error}) \leqslant d\delta$.

Proof. Whenever a rectangle R is put inside Error, $\mu_{R_0}(R) < \delta$. In any single layer of the execution tree, the nodes corresponding to calls of Decompose are associated with rectangles that are mutually disjoint. Thus, the total measure of all the rectangles that are included in Error in any layer is at most δ. There are at most d layers that have Decompose nodes that could lead to children that are included in Error. Thus, $\mu(\mathsf{Error}) \leqslant d\delta$. □

The main task is thus to upper bound bounding $\mu(\mathsf{Error}_d)$.

Bounding $\mu(\mathsf{Error}_d)$:. We begin with an important definition. Let w be the child associated with the i^{th} call of XDecompose or YDecompose inside the call of Decompose associated with some node v.

Definition 2.5.12 ($\Theta(w \mid v)$)**.** We define $\theta(w \mid v)$ to be the relative measure of the rectangle R inside Step 4 of Algorithm 2.5.3.1 just before the call of XDecompose or YDecompose associated with the node w.

$\theta(w \mid v)$ can be expressed in terms of μ as follows:

Lemma 2.5.13 (Measure at w Nodes)**.** *Let $w_1, w_2, \ldots,$ be the w-nodes that are children of v in the execution of the algorith. Then,*

$$\theta(w_i|v) = \sum_{j=i}^{c_v} \mu(w_j|v).$$

Proof. Let R_i be the rectangle that step 4 of Decompose works with just before the call of XDecompose or YDecompose that creates the node w_i. Each iteration of the while loop in step 4 of the algorithm further partition R. The total μ-measure of R_i can thus be obtained by summing up the measures of all the rectangles R_j that occur in Decompose at v at or after the call of XDecompose or YDecompose that creates w_i. □

Lemma 2.5.14. *Fix any vertex v and a child w of v,*

$$\mu(w|v) \geqslant 2^{-0.9b|S_w|} \cdot \theta(w|v).$$

Proof. Let R be the rectangle processed in the while loop inside the call of Decompose associated with v. Suppose w is created with the call of XDecompose or YDecompose for a choice of S such that $\mathbb{P}_{\mu_{|R}}[X_S = \alpha] \geqslant 2^{-0.9b|S|}$ (the other case can be dealt with analogously). Thus, $\mu(w \mid v) = \mu_{|R}(X_{|S=\alpha} \times Y) \geqslant 2^{-0.9b|S|}$.

But $\mu(w \mid v) = \mu(w \mid R)\mu(R \mid v)$. By Lemma 2.5.13, $\theta(w \mid v)$ is the relative measure of R inside $R(v)$. Thus, $\mu(w \mid v) \geqslant 2^{-0.9b|S|}\mu(R \mid v) = 2^{-0.9b|S|}\theta(w \mid v)$.

□

Next, we estimate $\mu(v)$ for every v that is a *bad* leaf i.e., a leaf at depth $2d + 1$.

Lemma 2.5.15. *Suppose $H_\infty(\mu) \geqslant 2nb - t$. Let $v_1, w_1, \ldots, v_d, w_d, v_{d+1}$ be a bad path. Then,*

$$\mu(v_{d+1}) \leqslant 2^{-1.1b|U_{v_{d+1}}|} \cdot 2^t \cdot \prod_{i=1}^{t} \frac{\mu(w_i|v_i)}{\theta(w_i|v_i)}$$

Proof. Let $s = |U_{v_{d+1}}|$; then, v_{d+1} has $2s$ blocks fixed. Using that $H_\infty(\mu) \geqslant 2nb - t$, we get that

$$\mu(v_{d+1}) \leqslant 2^{-2bs} \cdot 2^t.$$

On the other hand, we know that $\sum_{i=1}^{d} |S_{w_i}| = s$ and by Lemma 2.5.14 that

$$\mu(v_{d+1}) = \prod_{i=1}^{d} \mu(v_{i+1}|w_i)\mu(w_i|v_i) \geqslant \prod_{i=1}^{d} \mu(v_{i+1}|w_i) \cdot \theta(w_i|v_i) \cdot 2^{-0.9b|S_{w_i}|}.$$

The above two inequalities imply that,

$$\prod_{i=1}^{d} \mu(v_{i+1}|w_i)\theta(w_i|v_i) \leqslant 2^{-1.1bs} \cdot 2^t.$$

The claim now follows from the above inequality along with

$$\mu(v_{d+1}) = \prod_{i=1}^{d} \mu(v_{i+1}|w_i)\mu(w_i|v_i) = \left(\prod_{i=1}^{d} \mu(v_{i+1}|w_i)\theta(w_i|v_i)\right) \cdot \prod_{i=1}^{d} \frac{\mu(w_i|v_i)}{\theta(w_i|v_i)}.$$

□

Lemma 2.5.16. *Let $a_1, \ldots, a_N \in (0,1)$ be such that $\sum_{i=1}^{N} a_i = 1$ and for every $i < N$, $\sum_{j \geqslant i} a_j > \varepsilon$. Then,*

$$\sum_{j=1}^{N} \frac{a_j}{\sum_{i \geqslant j} a_i} \leqslant \lceil \log(1/\varepsilon) \rceil + 2.$$

Proof. For $i < N$, let $s_i = \sum_{j \geqslant i} a_j$; clearly, s_i is a decreasing sequence and $s_{N-1} > \varepsilon$. Let ℓ_i be the largest index such that $s_{\ell_i} = \sum_{j=\ell_i}^{N} a_j \geqslant \frac{1}{2^i}$.

By definition,

$$\sum_{j=\ell_i+1}^{\ell_{i+1}} a_j \leqslant \sum_{j=\ell_i+1}^{N} a_j < \frac{1}{2^i}$$

Let $t = \lceil \log(1/\varepsilon) \rceil$. Clearly, $\ell_t \geqslant N - 1$. Now, we have,

$$\begin{aligned}
\sum_{i=1}^{N-1} \frac{a_i}{s_i} &= \sum_{i=0}^{t} \sum_{j=\ell_i+1}^{\ell_{i+1}} \frac{a_j}{s_j} \\
&\leqslant \sum_{i=0}^{t} \frac{\sum_{j=\ell_i+1}^{\ell_{i+1}} a_j}{s_{\ell_{i+1}}} \\
&\leqslant \sum_{i=0}^{t} 1 = t + 1.
\end{aligned}$$

□

We have the following immediate consequence.

Corollary 2.5.17. *For every vertex v,*

$$\sum_{w \in C(v)} \frac{\mu(w|v)}{\theta(w|v)} \leqslant \lceil \log(1/\delta) \rceil + 2.$$

Lemma 2.5.18. *The sum over all leaves*

$$\sum_{\text{paths } v_1, w_1, \ldots, w_{\ell-1}, v_\ell} \prod_{i=1}^{\ell} 2^{-b|U_{v_\ell}|} \cdot \frac{\mu(w_i|v_i)}{\theta(w_i|v_i)} \leqslant \left(\lceil \log 1/\delta \rceil + 2 \right)^d$$

Proof. For each v corresponding to a Decompose call, define a probability distribution $\gamma(w|v)$ over $C(v)$ as follows.

$$\gamma(w|v) = \frac{\mu(w|v)}{\theta(w|v)} \cdot \left(\sum_{w \in C(v)} \frac{\mu(w|v)}{\theta(w|v)} \right)^{-1}.$$

For each w corresponding to a XDecompose or YDecompose call, define a uniform distribution over all its children. Then the above sampling procedure induces a probability distribution over the leaves of the execution tree - the probability of any leaf $v_{\ell+1}$ being given by $\prod_{i=1}^{\ell} 2^{-b|S_{w_i}|} \cdot \gamma(w_i|v_i) = 1$, where $v_1, w_1, \ldots, v_\ell, w_\ell, v_{\ell+1}$ is the path in the tree from the root that leads to $v_{\ell+1}$.

By our construction, the total probability under the above distribution of all the leaves is 1. Thus,

$$\sum_{\text{paths } v_1, w_1, \ldots, v_\ell w_\ell} \prod_{i=1}^{\ell} 2^{-b|S_{w_i}|} \cdot \gamma(w_i|v_i) = 1.$$

By Corollary 2.5.17, $\gamma(w_i|v_i) \geqslant \frac{\mu(w_i|v_i)}{\theta(w_i|v_i)} \cdot \frac{1}{\left(\lceil \log 1/\delta \rceil + 2\right)^d}$. Substituting this and using the fact that the length of any path is at most $2d$, the result follows. □

Lemma 2.5.19. *When the distribution μ has min-entropy at least $2n - t$,*

$$\sum_{v_{d+1} \in Bad} \mu(v_{d+1}) \leqslant 2^{-0.1bd} \cdot 2^t \cdot \left(\lceil \log 1/\delta \rceil + 2\right)^d.$$

Proof. The proof follows by using Lemmas 2.5.15 and Lemma 2.5.18:

$$\sum_{v_{d+1} \in Bad} \mu(v_{d+1}) \leqslant \sum_{v_{d+1} \in Bad} 2^{-1.1b|U_{v_{d+1}}|} \cdot 2^t \cdot \prod_{i=1}^{t} \frac{\mu(w_i|v_i)}{\theta(w_i|v_i)}$$

$$\leqslant 2^{-0.1bd} \cdot 2^t \left(\sum_{v_{d+1} \in Bad} 2^{-b|U_{v_{d+1}}|} \cdot \prod_{i=1}^{t} \frac{\mu(w_i|v_i)}{\theta(w_i|v_i)} \right) \leqslant 2^{-0.1bd} \cdot 2^t \cdot \left(\lceil \log 1/\delta \rceil + 2 \right)^d$$

□

Proof of Lemma 2.5.9. Follow from stringing together Lemmas 2.5.10 and 2.5.11 and 2.5.19.

□

Chapter 3

Sum of Squares Lower Bounds for Pairwise Independent CSPs

In this chapter[1], we prove that there is a constant c such that 2^{cn} time algorithm from the Sum-of-Squares hierarchy cannot beat the approximation guarantee of random assignment for any CSP on *pairwise independent* predicate. The results of this chapter were obtained in joint work with Boaz Barak and Siu On Chan [20].

We begin by formally stating the results. Towards this, we set up some definitions.

Definition 3.0.1 (Pseudo-expectation)**.** For every n and d, let $\mathcal{P}^n_d$ denote the linear space of n-variate real polynomials of degree at most d. A linear operator $\tilde{\mathbb{E}} : \mathcal{P}^n_d \to \mathbb{R}$ is a *degree-d pseudo-expectation operator* if it satisfies:

Normalization $\tilde{\mathbb{E}}[1] = 1$ where on the LHS 1 denotes the constant polynomial p such that $p(x) = 1$.

Positivity $\tilde{\mathbb{E}}[p^2] \geqslant 0$ for every $p \in \mathcal{P}^n_{d/2}$.

[1]The results of this chapter were published in the Proceedings of Symposium on Theory of Computing, STOC, 2015 in a paper titled *Sum of Squares Lower Bounds from Pairwise Independence* co-authored with Boaz Barak (Harvard U.) and Siu On Chan (Chinese University of Hong Kong). All co-authors contributed equally in producing all the results in the paper and are listed in alphabetical order in the proceedings.

For every polynomial $p \in \mathcal{P}^n_d$, we say that $\tilde{\mathbb{E}}$ satisfies the constraint $\{p = 0\}$ if $\tilde{\mathbb{E}}[pq] = 0$ for every $q \in \mathcal{P}^n_{d-\deg(P)}$.

The Sum-of-Squares hierarchy can be thought of as optimizing over pseudo-expectations; see the survey [28] and the references therein, as well as the lecture notes [17]. For notational convenience, we will use variables over $\{\pm 1\}$ instead of $\{0,1\}$. A *literal* is a function $f : \{\pm 1\}^n \to \{\pm 1\}$ such that $f(x) = x_i$ or $f(x) = -x_i$ for some i. If $C = (f_1, \ldots, f_k)$ is a k-tuple of literals then we denote by $C(x)$ the tuple $(f_1(x), \ldots, f_k(x))$. Our main result is the following:

Theorem 3.0.2 (Main Result)**.** *For every $k \in \mathbb{N}$, $\varepsilon > 0$ there exists $\delta = \delta(k) > 0$ such that for every sufficiently large $n \in \mathbb{N}$ there is a set $\mathcal{I} = \{C_1, \ldots, C_m\}$ of k-tuples of literals over $x_1, \ldots, x_n$ such that*

1. *For every $x \in \{\pm 1\}^n$, the distribution $\{C(x)\}$ where C is chosen at random in $\mathcal{I}$ is within ε statistical distance to the uniform distribution over $\{\pm 1\}^k$.*

2. *For every pairwise independent distribution μ over $\{\pm 1\}^k$, there exists a degree δn pseudo-expectation operator $\tilde{\mathbb{E}}$ over $\mathbb{R}^n$ satisfying the constraints $\{x_j^2 = 1\}_{j=1\ldots n}$ such that for every $C \in \mathcal{I}$ and $f : \{\pm 1\}^k \to \mathbb{R}$, $\tilde{\mathbb{E}}\, f(C(x)) = \mathbb{E}\, f(\mu)$.*

The following immediate corollary implies that predicates supporting pairwise independent distributions are approximation-resistant for $\Omega(n)$-degree SOS:

Corollary 3.0.3. *For every $\varepsilon > 0$ and $P : \{\pm 1\}^k \to \{0,1\}$, if there exists a pairwise independent distribution μ supported on $P^{-1}(1)$ then there exists $\delta > 0$ such that for all n there is a set $\mathcal{I} = \{C_1, \ldots, C_m\}$ of k-tuples of literals over $x_1, \ldots, x_n$ such that*

1. *For every* $x \in \{\pm 1\}^n$, $\mathbb{E}_{C \in \mathcal{I}} P(C(x)) \leqslant \frac{|P^{-1}(1)|}{2^k} + \varepsilon$.

2. *The value of the* δn*-degree Max-P SOS relaxation for the fraction of satisfiable constraints on the instance* $\mathcal{I}$ *is* 1.

Remark 3.0.4. The instance $\mathcal{I} = (C_1, \ldots, C_m)$ is actually obtained at random (with some pruning of a small fraction of the constraints, or alternatively, with some loss in the "perfect completeness" condition). Thus our results can also be thought as giving some evidence to a conjecture of Barak, Kindler and Steurer [26] that no polynomial-time algorithm (including in particular the SOS algorithm) can beat the basic semidefinite program on approximating random CSP instances.

Throughout this paper we restrict ourselves to the *Boolean* case, and do not consider extensions to a larger alphabet, though our methods may be useful in this case as well.

3.1 Related works

Grigoriev [59] proved in 1999 that (in the language of this paper) 3XOR is approximation resistant for the degree $\Omega(n)$ Sum-of-Squares hierarchy. Grigoriev's work in fact predated the papers of Parrilo [90] and Lasserre [80] proposing the SOS hierarchy, and so he used the different (but equivalent) language of Positivstellensatz Calculus proofs. (Also, as far we know, he did not note that these proofs can be efficiently found via a semidefinite program.) Grigoriev's result was rediscovered in 2008 by Schoenebeck [95], who also noted that it implies approximation resistance for 3SAT and some other CSPs as well. Tulsiani [103] (see also Chan [37]) further generalized these results and in particular

showed that every predicate that contains a pairwise independent subgroup is approximation resistant for $\Omega(n)$-degree SOS. Both Tulsiani and Schoenebeck follow Grigoriev's technique of reducing SOS lower bounds to resolution width lower bounds. As far as we know, no other SOS integrality gaps for approximating CSPs were known, and there are very few SOS lower bounds in general, most notably Grigoriev's lower bound for knapsack [58] and the very recent result by Meka, Potechin and Wigderson for the planted clique problem (personal communication).

Arora, Bollobás, Lovász and Tourlakis [10] obtained integrality gaps for the Lovász-Schrijver linear programming hierarchy for Vertex Cover. Schoenebeck, Trevisan and Tulsiani [97] showed that Max-Cut is approximation resistant for $\Omega(n)$ levels of the Lovász-Schrijver hierarchy, and these results have been strengthened to the stronger Sherali-Adams hierarchy [42, 39]. The famous Goemans-Williamson algorithm [54] shows that Max-Cut is *not* approximation resistant for even the degree 2 SOS hierarchy, further underscoring the difference between these relaxations.

Perhaps closest to our work are the papers of Benabbas, Georgiou, Magen, and Tulsiani [30] who showed that predicates containing a pairwise independent distribution are approximation resistant for $\Omega(n)$ rounds of the Sherali Adams hierarchy, even when one adds the degree 2 SOS constraints. Indeed, our pseudo-distribution agrees with theirs, though we describe it somewhat differently, and most importantly, need a completely different argument to show that it is positive semi-definite. Our work is also inspired by the pseudo-expectation view of the SOS hierarchy as advocated in the papers [19, 23].

3.2 Overview of our proof

To prove Theorem 3.0.2, we need to show that given any pairwise independent distribution μ over $\{\pm 1\}^k$, one can come up with $\mathcal{I}$, a collection of tuples $\{C_1, \ldots, C_m\}$ of literals and a pseudo-expectation operator $\tilde{\mathbb{E}}$ that "pretends" to be the expectation of a valid distribution whose projection on to any C_i is μ. In fact, our choices for both $\mathcal{I}$ and $\tilde{\mathbb{E}}$ will not be novel and follow prior works in this area. For $\mathcal{I}$, as mentioned, we will simply use a random set of tuples (or more accurately, a set corresponding to a hypergraph with sufficiently strong expansion properties), as was done by previous works dealing with weaker hierarchies [30, 105, 89]. It turns out that given this choice, the pseudo-expectation $\tilde{\mathbb{E}}$ is essentially "forced", and again, we use the same pseudo-expectation used in prior works such as [30], though we describe it slightly differently. This pseudo-expectation corresponds in some sense to the "maximum entropy distribution" conditioned on satisfying our constraints (though of course it is not an actual distribution but only a *pseudo-distribution* in the sense of [28]). Those prior works have shown that for every set S of $o(n)$ variables, there is a distribution ν_S over the variables in S that agrees with $\tilde{\mathbb{E}}$. The main difference is that we prove that for some $d = \Omega(n)$, $\tilde{\mathbb{E}}$ is a valid degree-d pseudo-expectation operator, that is, it satisfies the non-negativity / positive semidefiniteness condition $\tilde{\mathbb{E}}[p^2] \geqslant 0$ for every polynomial $p \leqslant d/2$. This is a more "global" property, as the polynomial p might depend on all n variables, which makes it more challenging to prove.

Our approach is to essentially diagonalize $\tilde{\mathbb{E}}$. That is, we will show an explicit construction of polynomials $\tilde{\chi}_1, \ldots, \tilde{\chi}_M \in \mathcal{P}^n_{d/2}$ which we call *local orthogonal functions* such that **(1)** $\{\tilde{\chi}_i\}_{i=1}^M$ spans the space $\mathcal{P}^n_{d/2}$, **(2)** $\tilde{\mathbb{E}}[\tilde{\chi}_i \tilde{\chi}_j] = 0$ for all $i \neq j$ and **(3)** $\tilde{\mathbb{E}}[\tilde{\chi}_i^2] \geqslant 0$

for all i. The existence of these polynomials immediately implies the property we need, as, by representing every polynomial p as $p = \sum_i p_i \tilde{\chi}_i$, we see that

$$\tilde{\mathbb{E}}[p^2] = \sum_{i,j} p_i p_j \, \tilde{\mathbb{E}}[\tilde{\chi}_i \tilde{\chi}_j] = \sum_i p_i^2 \, \tilde{\mathbb{E}}[\tilde{\chi}_i^2] \geqslant 0 \; .$$

We now review the construction of the instance, as well as the pseudo-expectation operator, and then discuss how we come up with these local orthogonal functions. As mentioned above, our instance $\mathcal{I} = (C_1, \ldots, C_m)$ will simply be a random instance, which we think of as a k-uniform hypergraph with m hyperedges $C_1, \ldots, C_m$. After some pruning we can assume this hypergraph has girth $\Omega(\log n)$.[2] By a simple Chernoff + union bound argument, if $m > cn$ for a sufficiently large constant c then for every assignment $x \in \{\pm 1\}^n$, the induced distribution $\{C_i(x)\}_{i \sim [m]}$ will be ε-close to the uniform distribution. For this informal overview, suppose that we merely want to establish the existence of a degree d pseudo-expectation operator for some large constant d. Note that this means that sets of at most d (or even 2^d) variables form a *forest* (i.e. disjoint collection of trees) in this hypergraph.

We now describe the pseudo-expectation operator $\tilde{\mathbb{E}}$, which in some sense is almost "forced" as the only natural operator for this instance. (As mentioned, this part is not novel and the same operator was used by works such as [30]; however we describe it somewhat differently.) We construct $\tilde{\mathbb{E}}$ by defining for every set S of at most d variables a distribution ν_S over $\{\pm 1\}^S$ such that **(1)** for every clause C contained in S, the projection

[2] If we don't prune these clauses then our proof guarantees that for $1 - o(1)$ fraction of the clauses we get the marginal distribution to be μ. It is possible that this can be upgraded to all of the clauses at the expense of some additional complication, but we have not checked whether or not that's the case.

of ν_S to C equals μ and **(2)** the distributions are *locally consistent* in the sense that if $S \subseteq U$ then the projection of ν_U to S equals ν_S. The definition of ν_S is very simple. First, say for the purposes of this informal overview that a set S is *closed* if every clause C in $\mathcal{I}$ is either completely contained in S or intersects it in at most a single variable. If S of size $O(d)$ is closed and connected (as a subgraph of $\mathcal{I}$) then it is a *tree* in the hypergraph $\mathcal{I}$. In this case, we define the distribution ν_S as follows: to sample x from ν_S we pick an arbitrary clause $C \subseteq S$ and sample its variables according to μ. We then continue down the tree, sampling the variables of all the clauses that intersect with C, and so on. It is not hard to show that because of pairwise independence (and in fact simply because every marginal is uniform) this process will always yield the same distribution regardless of the traversal order, and the probability of $x \in \{\pm 1\}^S$ to be sampled under this distribution will be proportional to $\prod_{C \subseteq S} \mathbb{P}[\mu = C(x)]$. If a set S is closed but not connected then the distribution ν_S is obtained by making independent choices for each of the connected components of S. For a general (not necessarily closed) set S, we define the *closure* of S, denoted by $\mathsf{cl}(S)$, to be the *minimal* closed superset of S (this is well defined; one can show that intersections of closed sets are closed and thus, *the* minimal closed set is the intersection of all closed sets containing S). A fairly simple argument using the girth condition can be used to argue that $|\mathsf{cl}(S)| \leqslant O(|S|)$ for every $|S| \leqslant d$. We then define ν_S to be the distribution obtained by projecting the distribution $\nu_{\mathsf{cl}(S)}$ to S. The collection of local distributions so obtained satisfies **(1)** by construction, and it is not hard to show that it satisfies **(2)** as well. Since all polynomials of degree at most d are spanned by the set of polynomials $\{\chi_S\}_{|S| \leqslant d}$ (which we will call the *characters*) where $\chi_S(x) = \prod_{i \in S} x_i$, to define the pseudo-expectation operator it suffices to define $\tilde{\mathbb{E}}[\chi_S]$ for every $|S| \leqslant d$. We simply define $\tilde{\mathbb{E}}[\chi_S]$ to be $\mathbb{E}_{x \sim \nu_S}[\chi_S(x)]$.

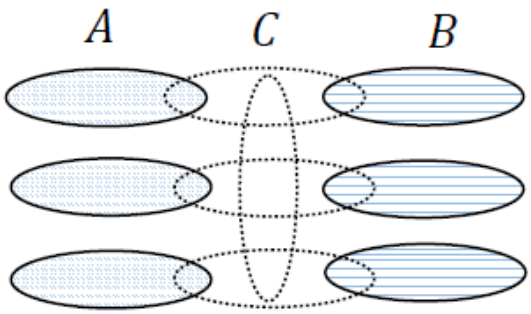


Figure 3.1: In this example, even though both A and B are collections of disjoint clauses and hence are "closed" under our definition, their distributions could be correlated due to the existence of the set C.

We now describe how we come up with the functions $\tilde{\chi}_1, \ldots, \tilde{\chi}_M$. Intuitively, we would like to come up with these functions via a Gram-Schmidt like process. That is, we fix some ordering $A_1 \prec \ldots \prec A_M$ of the $M = \binom{n}{\leqslant d}$ sets of size at most d, and define χ_i to be χ_{A_i}. Now, we would want to define $\tilde{\chi}_i$ to be the component orthogonal to the span of $\chi_1, \ldots, \chi_{i-1}$ where we define orthogonality using $\tilde{\mathbb{E}}$ as an inner product. We would then get that $\tilde{\mathbb{E}}\, \tilde{\chi}_i \chi_j = 0$ for all $j < i$, which would imply that $\tilde{\mathbb{E}}\, \tilde{\chi}_i \tilde{\chi}_j = 0$ for all $i \neq j$ (as $\tilde{\chi}_j$ is spanned by $\chi_1, \ldots, \chi_j$). However, this is of course circular reasoning, since we cannot assume that $\tilde{\mathbb{E}}$ is positive semidefinite (and hence a valid inner product) since this is exactly what we are trying to prove!

However, because we know that on every small set U, $\tilde{\mathbb{E}}$ agrees with an actual expectation operator (the one associated with the *actual* distribution ν_U), we do know that it is psd when it is restricted to this small set U. Therefore, if for some reason when we do this Gram-Schmidt process and express $\tilde{\chi}_i$ as some linear combination $\sum_{j \leqslant i} \alpha_j \chi_j$, we get lucky and this linear combination happens to be extremely *sparse* then we can actually

carry through the argument described above. Specifically, it turns out that it suffices for the set $U = \cup\{A_j \mid \alpha_j \neq 0\}$ to be sufficiently small so that $\tilde{\mathbb{E}}$ is a valid inner product on $U \cup A_i$. However a priori, this hope seems dubious, since the Gram-Schmidt process is very sequential, and we need to do it for $\binom{n}{\leqslant d}$ steps. It seems quite possible that we would create *long distance correlations* in the process, whereby we would end up needing to express $\tilde{\chi}_i$ using many χ_j's for sets A_j that are quite far from A_i. (See Figure 3.1 for one example of a correlation that could arise between two disjoint collection of clauses A and B.)

Nevertheless, we show that we are in fact able to choose a tailor-made ordering of the sets so that this hope is (essentially) materialized. An important observation that comes to our aid here is that our local distributions, intuitively speaking, satisfy: if two sets A and B are sufficiently far apart in the hypergraph $\mathcal{I}$, then the distribution $\nu_{A\cup B}$ is obtained by taking the product of the independent distributions ν_A and ν_B. We use this observation to argue that, if we choose the ordering on the sets in $\binom{[n]}{d}$ in the right way, then, when we express $\tilde{\chi}_i$ as a linear combination of the functions χ_j for $j < i$, we only use j's such that A_j is contained in a certain (carefully defined) small "ball" in the hypergraph around the set A_i. The crucial result that we need here is to show that whenever there is a dependence between the local distribution on some set A and the local distribution on some set B that came *before* A in our order, then, either B is contained in this "ball" around A, or the correlation between A and B is completely "explained" by the intersection of the closure of B with this ball, in the sense that conditioned on any assignment to the variables in the intersection, the local distributions on A and B are independent. This will allow us to argue that we don't need to use χ_B to express χ_{A_i} but can restrict ourselves to characters

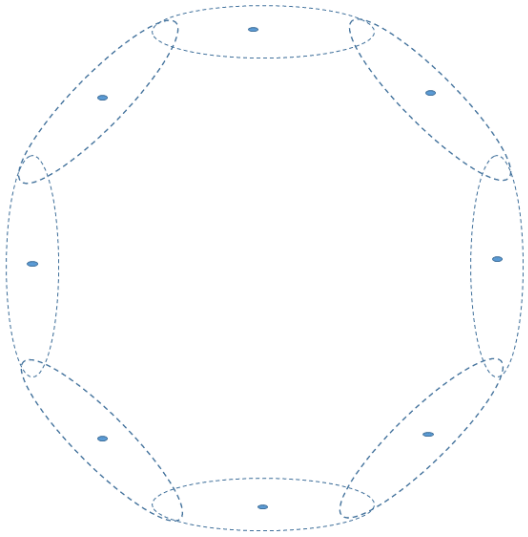


Figure 3.2: In this example, the solid dots are variables and no clause contains any two of them, but the local distribution on the variables might not be uniform since the constraints of the cycle can create a dependency.

contained in that ball. Moreover, and crucially, we will show that our ordering has the property that all the characters we will need to use must have come before A as well.

Handling $\Omega(n)$ rounds.. The above overview can be converted into a full proof with some care when $d = o(\log(n))$ by exploiting the acyclicity of all subgraphs involved. Extending to $d = \Omega(n)$, however, introduces additional subtleties. When d exceeds $\Omega(\log(n))$, subgraphs induced by d vertices of $\mathcal{I}$ can have cycles. An immediate effect of this is that the the definition of a closed set that we gave before no longer yields consistent local distributions on any collection of d variables. An example of a problem that arises when cycles can exist on a set of vertices is illustrated in Figure 3.2. To fix this, we define a stronger notion of closed set S that guarantees that all paths of length at most 3 between

any two vertices in S are completely contained inside S. This notion of closures differs from the one that Benabbas et. al. [30] use. An appeal to the expansion property of $\mathcal{I}$ (instead of high girth as before) can be used to show that the closure of a set S is at most a constant factor larger than $|S|$. Similarly, as before, we need to show that there exists a (suitably defined) ball, $Ball(A)$ around any set A of variables (of size at most d) such that the correlations with any other set B of size at most d are "captured" by the intersection of $Ball(A)$ and B. This needs a more careful argument. In particular, the correlations (even in the low girth case) are actually not necessarily captured by the intersection of $Ball(A)$ with B, but rather with some set B' that is related, but not identical to B. However, the crucial property that we require is that the set $B_{in} = Ball(A) \cap B'$ satisfies **(1)** if B came before A in the ordering, then so will B_{in} and **(2)** $|B_{in}| + |B \setminus Ball(A)| \leqslant |B|$. This second property is more complicated to prove in the case where $|B|$ can be much larger than the girth bound, but turns out to hold there as well. The bottom line is that with additional care however, the high level picture provided by this overview can indeed be implemented and we give a full analysis based on the local Gram-Schmidt like process in Section 3.6.

3.3 Preliminaries

We collect some standard definitions and notation here. A (k, n)-*instance* is a k-uniform hypergraph $\mathcal{I} = \{C_1, \ldots, C_m\}$ over $[n]$ so that every hyperedge (also known as a *clause*) $C = (i_1, \ldots, i_k) \in \mathcal{I}$ is labeled by a string $\sigma = \sigma^C \in \{\pm 1\}^k$. We identify a clause C with the function that maps $x \in \{\pm 1\}^n$ to $y_1, \ldots, y_k$ where $y_j = \sigma_j x_{i_j}$. We will sometimes also consider C as a tuple of the *literals* $(\sigma_{i_1} x_{i_1}, \ldots, \sigma_{i_k} x_{i_k})$. We write $V(C)$ for the variables involved in (or *covered by*) a clause C and similarly for $V \subseteq [n]$ we write $\mathbb{C}(V)$ for the set of

all clauses C such that $V(C) \subseteq V$. For any $x \in \{-1,1\}^n$, we write x_A to denote the tuple of coordinates in the subset $A \subseteq [n]$. If $x \in \{-1,1\}^A$ and $y \in \{-1,1\}^B$ for disjoint sets A and B, we will write $x \circ y$ for the string in $\{-1,1\}^{A \cup B}$ that projects to x for coordinates in A and to y for coordinates in B.

Unless explicitly mentioned, the base of all logarithms appearing in the paper is assumed to be 2. We consider the arity of our tuples k to be a constant and so O notation may hide the dependence on k.

We now define some standard ideas in the context of hypergraphs.

Definition 3.3.1. Let G be a hypergraph. G is said to be a *path* if its hyperedges can be ordered into a sequence $C_1, C_2, \ldots, C_\ell$ such that for each $2 \leqslant i \leqslant \ell$, $C_i \cap C_{i-1} \neq \emptyset$ and $C_i \cap C_j = \emptyset$ for every $|i-j| > 1$. G is said to be a *cycle* if it has at least two hyperedges, and there is a cyclic ordering of its hyperedges $C_0, C_1, \ldots, C_{\ell-1}$, and there are distinct vertices $v_0, \ldots, v_{\ell-1}$ with $v_i \in C_i \cap C_{(i+1) \bmod \ell}$ for all i. G is said to be a *forest* if it does not contain any cycle. A forest is a *tree* if it is *connected* (i.e. for every two distinct vertices u and v, there is a path $C_1, \ldots, C_\ell$ such that $u \in C_1$ and $v \in C_\ell$).

The *degree* of G is the maximum number of hyperedges that intersect with any given hyperedge in G. The length of the shortest cycle in G is said to be the *girth* of G. For any vertices u, v of a hypergraph G, we define the *distance*, $\mathrm{dist}(u,v)$ of u, v in G as the minimum number of hyperedges in any path that joins u and v in G. For S, T, subsets of vertices, we define $\mathrm{dist}(S,T) \stackrel{\text{def}}{=} \min_{s \in S, t \in T} \mathrm{dist}(s,t)$.

Next, we define the notion of expansion in a k-uniform hypergraph G:

Definition 3.3.2 (Coefficient of Expansion)**.** A k-uniform constraint hypergraph G is said to be (r, β)-expanding if any collection $\mathbb{C}$ of at most r hyperedges of G cover at least $(k-1-\beta)|\mathbb{C}|$ vertices of G, i.e. $|\{v \mid \exists C \in \mathbb{C}, v \in C\}| \geqslant (k-1-\beta)|\mathbb{C}|$. We call β, the *coefficient of expansion* of G.

Let $\mathcal{I}$ be a (k, n) instance. We now describe the properties of the (k, n) instances that we need and give a construction for them in Section B.2 of the Appendix by taking a random instance and removing a few clauses. Specifically, we show the existence of *nice* instances, the ones that satisfy the properties described in the lemma below:

Lemma 3.3.3. *Fix $1 > \varepsilon, \delta \geqslant 0$ and $\gamma \geqslant e^k k^2$. Then, there exists a k-uniform constraint hypergraph G with γn edges such that for $\eta = (1/\gamma^2)^{2/\delta}$, $1/\tau = 4\log_2(\gamma k^2)$, G:*

1. *is $(\eta n, \delta)$-expanding,*
2. *has girth $\mathfrak{g} \geqslant \tau \log(n)$*

We will use this lemma with any given ε (the soundness slack), $\delta = \frac{1}{200}$ and $\gamma = e^k k^2/\varepsilon^2$. We will call the instances that satisfy the conditions of the lemma above as *nice*.

For such instances, it is also easy to prove the soundness part (part (i)) of Theorem 3.0.2 (see Section B.2.1 of the Appendix) which we record in the following lemma.

Lemma 3.3.4. *For every $\varepsilon > 0$ and k, if n is sufficiently large then there exists a nice (k, n)-instance $\mathcal{I}$ with the property that for every $x \in \{\pm 1\}^n$, the distribution $\{C(x)\}_{C \in \mathcal{I}}$ is ε-close in total variation distance to the uniform distribution on $\{\pm 1\}^k$.*

3.4 Closed sets, and the definition of the pseudo-expectation

Throughout the rest of this paper we fix $\mathcal{I} = (C_1, \ldots, C_m)$ to be a nice (k, n) instance with coefficient of expansion β. Thus whenever we mention edges, paths, or clauses, they will always be with respect to the hypergraph $\mathcal{I}$. In this section, we define a linear operator $\tilde{\mathbb{E}}$ on P^n_s, the linear space of multilinear polynomials on $\mathbb{R}^n$ of degree at most $s = \frac{\eta n}{6}$. We will ensure that the $\tilde{\mathbb{E}}$ so defined will satisfy $\tilde{\mathbb{E}}[f(C(x))] = \mathbb{E}[f(\mu)]$ for every clause $C \in \mathcal{I}$ and function $f : \{\pm 1\}^k \to \mathbb{R}$. In the next section, we will show that the $\tilde{\mathbb{E}}$ we define here is in fact a pseudo-expectation operator on P^n_d for $d = \frac{\eta n}{10000k}$ and thus obtain our main result. The $\tilde{\mathbb{E}}$ operator we use was defined in previous works such as Benabbas et. al. [30] and later also used by Tulsiani and Worah [105] to study weaker LP/SDP hierarchies. Here, we describe a construction of the same operator in a slightly different way so as to help us in the proof of our main result.

To define $\tilde{\mathbb{E}}$, it is enough to define $\tilde{\mathbb{E}}[\chi_S]$ for characters χ_S for each $S \subseteq [n], |S| \leqslant s$, as one can then extend $\tilde{\mathbb{E}}$ linearly to all of P^n_s. To do this, we define a probability distribution ν_X for every $X \subseteq [n]$, such that $|X| \leqslant s$, and then set $\tilde{\mathbb{E}}[\chi_S]$ to be the expectation of χ_S under ν_S.

3.4.1 Closures

We first define the concept of closed sets that is central to our argument.

Definition 3.4.1 (Closure and closed sets)**.** For every $R \geqslant 1$, a set $A \subseteq [n]$ is *R-closed* if for every $v, v' \in A$, any path of length at most R between v and v' is contained in A. We say that A is *closed* if it is 3-closed.

We define the *R-closure of A*, denoted by $\mathsf{cl}_R(A)$, to be the intersection of all sets B such that $A \subseteq B$ and B is R-closed. The *closure of A*, denoted by $\mathsf{cl}(A)$, is the 3-closure of A.

Remark 3.4.2. Readers familiar with the definition of closure (or advice set) in the work of [30] or [105] will find the definition of closure above slightly different. The main difference is that our definition allows us to have some nice properties such as uniqueness and that the intersection of two closed sets is closed, which are very helpful for our proof. We stress however that the actual pseudo-expectation is the same as that of those works.

Next, we give a constructive definition of closure of a set.

Lemma 3.4.3. *Given $S \subseteq [n]$ and any $R < \min\{\mathfrak{g}/2, \frac{1}{2\beta}\}$, the* R-closure *of S can be obtained by the following procedure run on S: Set $A := \emptyset$. For every $v, v' \in V(A) \cup S$ such that there is a path of length at most R between v and v' in $\mathcal{I}$ not contained in A, add every clause in the path to A. Output $V(A) \cup S$.*

Proof. Observe that the procedure terminates as there are only finitely many clauses. Further, the output is closed by virtue of the termination of the procedure. By induction on the time at which a path is added in the procedure, it is easy to show that every closed set containing S must contain the path. Thus, $V(A)$ is a closed set containing A and every clause C such that $V(C) \subseteq V(A)$ satisfies $V(C) \subseteq \mathsf{cl}_R(S)$. The lemma now follows by the minimality of $\mathsf{cl}_R(S)$. □

Next, we bound the size of $\mathsf{cl}_R(S)$.

Lemma 3.4.4. *For any $R < \min\{\mathfrak{g}/2, \frac{1}{2\beta}\}$ and $S \subseteq [n]$ such that $|S| \leqslant \frac{\eta n}{10R}$. Then, $|\mathbb{C}(\mathsf{cl}_R(S))| \leqslant 2R|S|$ and $|\mathsf{cl}_R(S)| \leqslant 2Rk|S|$.*

Proof. Consider the procedure described in Lemma 3.4.3. Let $S^{iso} \subseteq \mathsf{cl}_R(S)$ be the isolated vertices in $\mathsf{cl}_R(S)$. Observe that one cannot add any isolated vertices in the procedure and thus $S^{iso} \subseteq S$. Define $S' = S \setminus S^{iso}$. Then, $\mathsf{cl}_R(S) = \mathsf{cl}_R(S') \cup S^{iso}$.

If the process terminates before adding a total of $q = \frac{|S'|}{\frac{1}{R}-\beta}$ clauses, then there's nothing to prove, since $|S'| \leqslant |S| \leqslant \frac{\eta n}{10R}$ yields that $q \leqslant \frac{\eta n}{5}$. Thus, suppose, for the sake of a contradiction, that the procedure adds $> q$ clauses and let i^{th} round of the procedure be the first round where the number of clauses added exceeds q.

Let $\mathbb{C}_i$ be the set of clauses added in the procedure till the i^{th} round and let S'_i be the set of variables obtained by taking the union of variables covered by the clauses added and S'. Further, suppose that the i^{th} round adds q_i clauses. Then, $|\mathbb{C}_i| \leqslant q + q_i < \eta n$ and thus, $\mathbb{C}_i$ must satisfy the expansion requirement: $|V(\mathbb{C}_i)| \geqslant (q + q_i)(k - 1 - \beta)$. On the other hand, any new path of length $j \leqslant R$ added in a round adds at most $jk - (j-1) - 2$ new vertices. Thus, on an average, every one of the at most j new clauses added in any round of the procedure contribute at most: $k - 1 - 1/j \leqslant k - 1 - 1/R$ new vertices. Thus, $|S'_i| \leqslant |S'| + (q + q_i)(k - 1 - 1/R)$.

Now,

$$(q + q_i)(k - 1 - \beta) \leqslant |V(\mathbb{C}_i)| \leqslant |S'_i| \leqslant |S'| + (q + q_i) \cdot (k - 1 - 1/R).$$

This yields that $|S'| \geqslant (q + q_i) \cdot (1/R - \beta) > |S'|$ using that $q = \frac{|S'|}{\frac{1}{R}-\beta}$. This is a contradiction.

The size claimed in the lemma now follows by observing that $\frac{1}{R} - \beta \geqslant \frac{1}{2R}$ and that every clause contributes at most k new variables. □

The following lemma summarizes the simple properties of the closures defined here.

Lemma 3.4.5 (Simple Properties of Closures)**.** *1. For any $R < \mathfrak{g}/2$, if A and B are R-closed and then so is $A \cap B$.*

2. If $A \subseteq B$ then $\mathsf{cl}_R(A) \subseteq \mathsf{cl}_R(B)$.

3. Every connected component of $\mathsf{cl}_R(A)$ of size $\geqslant 2$ intersects A in at least two elements.

4. Let $A = A_1 \cup A_2 \cup \ldots A_m$. Then, $\mathsf{cl}(A) = \mathsf{cl}(\cup_{i=1}^{m} \mathsf{cl}(A_i))$.

Proof. 1. If there are two vertices v, v' in $A \cap B$ such that $\mathrm{dist}(v, v') \leqslant R$, then since both A and B are closed, both of them should contain the unique (since $R < \mathfrak{g}/2$) path between them.

2. By definition, $\mathsf{cl}_R(B)$ is an R-closed set containing $B \supseteq A$ and hence if $\mathsf{cl}_R(A) \nsubseteq \mathsf{cl}_R(B)$ then $\mathsf{cl}_R(A) \cap \mathsf{cl}_R(B)$ would be an even smaller R-closed set that contains A, contradicting the minimality of $\mathsf{cl}_R(A)$.

3. Suppose otherwise that there is some connected component S of $\mathsf{cl}_R(A)$ with $|S| \geqslant 2$ intersecting A with at most one element $\{x\}$, then we claim that $B = (\mathsf{cl}_R(A) \setminus S) \cup \{x\}$ is an R-closed set containing A. Clearly, $B \supseteq A$. Now suppose for the sake of contradiction that there were two vertices $v \neq v'$ of distance at most R in B whose path is not in B. Then since $B \subseteq \mathsf{cl}_R(A)$ and $\mathsf{cl}_R(A)$ is R-closed, the path between v

and v' must have had a vertex $u \in S \setminus \{x\}$. But since one of v or v' must be different than x (say v'), we get by contradiction that v' was connected to S in $\mathsf{cl}_R(A)$.

4. Let $B = \mathsf{cl}(\cup_{i=1}^{m} \mathsf{cl}(A_i))$. Since $\mathsf{cl}(A)$ is closed and contains $\cup_{i=1}^{m} A_i$, $B \subseteq \mathsf{cl}(A)$. If $B \neq \mathsf{cl}(A)$, then, $B \supseteq \cup_{i=1}^{m} A_i$ and is closed contradicting the minimality of $\mathsf{cl}(A)$. □

3.4.2 Definition of $\tilde{\mathbb{E}}$

Using the closures defined above, we define a local probability distribution on all closed sets and use it to define $\tilde{\mathbb{E}}$. Let $C = (v_1, v_2, \ldots, v_k)$, where, each v_j is the literal $\sigma_j x_{i_j}$ for some $\sigma_j \in \{\pm 1\}$. The distribution μ_C simply assigns to $x \in \{\pm 1\}^n$ the probability $\mu(\sigma_1 x_{i_1}, \ldots, \sigma_k x_{i_k})$ (i.e., the probability that $C(x) = a$ under μ_C is set to $\mu(a)$ for every $a \in \{\pm 1\}^k$).

The definition and the proof of consistency of the local distribution we define were shown by Benabbas et. al. [30] for the weaker notion of closures they used (in order to define linear round solutions in the Sherali Adams hierarchy). The argument for our notion of closure is similar but we include it here for the sake of completeness.

For every set $S \subseteq [n]$, $|S| \leqslant d$, let $\mathsf{cl}(S)$ be the closure of S and suppose I_S is the set of isolated variables in $\mathsf{cl}(S)$. Define $\mathbb{C}(\mathsf{cl}(S))$ be all clauses C such that $V(C) \subseteq \mathsf{cl}(S)$. Then, we set:

$$\nu_{\mathsf{cl}(S)}(x) = Z_{\mathsf{cl}(S)} \cdot \Pi_{C \in \mathbb{C}(\mathsf{cl}(S))} \mu_C(x_C) \tag{3.4.1}$$

where x_C the projection of x on to the coordinates in $V(C)$, and $Z_{\mathsf{cl}(S)} = 2^{k|\mathbb{C}(\mathsf{cl}(S))| - |\mathsf{cl}(S)|}$ ($\geqslant 1$). Observe that the above expression tells us that the marginal distribution of $\nu_{cl(S)}$ over I_S is uniform. We extend the notation above and write ν_T for the marginal of $\nu_{\mathsf{cl}(T)}$ on

variables in T.

We now show that $\nu_{\mathsf{cl}(S)}$ defined above is indeed a probability distribution over $\mathsf{cl}(S)$.

Lemma 3.4.6. *Let A and B be closed sets such that $A \subseteq B$ and $|\mathbb{C}(B)| \leqslant \eta n$. Then,*

1. *ν_A is a valid probability distribution: $\sum_{x \in \{-1,1\}^A} \nu_A(x) = 1$.*
2. *ν is locally consistent: for every $x \in \{-1,1\}^S$, $\nu_A(x) = \sum_{y \in \{-1,1\}^{B \setminus A}} \nu_B(x \circ y)$.*

The following claim that we record as a lemma will be useful in the proof.

Lemma 3.4.7. *There exists an ordering $C_1, C_2, \ldots, C_r$ of clauses in $\mathbb{C}_{A,B}$ and a partition of $B \setminus A$ into sets $F_1 \subseteq V(C_1), F_2 \subseteq V(C_2), \ldots, F_r \subseteq V(C_r)$ such that for every $j \leqslant r$, $|F_j| \geqslant k - 2$ and $F_j \cap \left(\cup_{i>j} V(C_i)\right) = \emptyset$.*

We first complete the proof of Lemma 3.4.6 and then prove Lemma 3.4.7.

Proof of Lemma 3.4.6. Let $Z_A = 2^{-|A|+k|\mathbb{C}(A)|}$ and $Z_B = 2^{-|B|+k|\mathbb{C}(B)|}$. Let $\mathbb{C}_{A,B} = \mathbb{C}(B) \setminus \mathbb{C}(A)$. Using (3.4.1), we have:

$$\begin{aligned}\sum_{y \in \{-1,1\}^{B \setminus A}} \nu_B(x \circ y) &= Z_B \cdot \Pi_{C \in \mathbb{C}(B)} \mu_C(x_C \circ y_C) \\ &= Z_B \cdot \Pi_{C \in \mathbb{C}(A)} \mu_C(x_C) \cdot \Pi_{C \in \mathbb{C}_{A,B}} \mu_C(x_C \circ y_C)\end{aligned}$$

To simplify notation, we will write μ_i for μ_{C_i} and x^i for $x_{V(C_i)}$ where $x \in \{-1,1\}^n$. We have, using the ordering given by Lemma 3.4.7. Then,

$$\sum_{y\in\{-1,1\}^{B\setminus A}} \nu_B(x \circ y) = Z_B \sum_{y\in\{-1,1\}^{B\setminus A}} \cdot \Pi_{C\in\mathbb{C}(A)}\mu_C(x_C) \cdot \Pi_{C\in\mathbb{C}_{A,B}}\mu_C(x_C \circ y_C)$$

$$= Z_B \sum_{y\in\{-1,1\}^{B\setminus A}} \Pi_{C\in\mathbb{C}(A)}\mu_C(x_C) \cdot \Pi_{i=1}^{r}\mu_i(x^i \circ y^i)$$

Using the partition $F_1, F_2, \ldots, F_r$

$$= Z_B \Pi_{C\in\mathbb{C}(A)}\mu_C(x_C) \cdot \sum_{\alpha_r\in\{-1,1\}^{F_r}} \mu_r(\zeta_r \circ \alpha_r) \cdots \sum_{\alpha_1\in\{-1,1\}^{F_1}} \mu_1(\zeta_1 \circ \alpha_1)$$

where ζ_r is the value for the variables in $V(C_r) \setminus F_r$.

Using that $|F_r| \geqslant k-2$ and pairwise independence of μ

$$= Z_B \Pi_{C\in\mathbb{C}(A)}\mu_C(x_C) \cdot \sum_{\alpha_r\in\{-1,1\}^{F_r}} \mu_r(\zeta_r \circ \alpha_r) \cdots \sum_{\alpha_2\in\{-1,1\}^{F_2}} \mu_2(\zeta_2 \circ \alpha_2) \cdot 2^{-|V(C_r)\setminus F_r|}$$

Continuing similarly for $2, 3, \ldots, r$

$$= Z_B \Pi_{C\in\mathbb{C}(A)}\mu_C(x_C) \cdot 2^{-\sum_{i=1}^{r}|V(C_r)\setminus F_r|}.$$

Now, $\sum_{i=1}^{r}|V(C_r) \setminus F_r| = kr - |B \setminus A|$. Further, $-|B| + k|\mathbb{C}(B)| - kr + |B \setminus A| = -|A| + k|\mathbb{C}(A)|$. Thus, $Z_B \cdot 2^{-\sum_{i=1}^{r}|V(C_r)\setminus F_r|} = Z_A$ completing the proof. □

We now complete the proof of Lemma 3.4.7.

Proof of Lemma 3.4.7. For every $C \in \mathbb{C}_{A,B}$ define $\Gamma(C) = \{v \in V(C) \mid \forall C' \neq C \in \mathbb{C}_{A,B}, v \notin V(C')\}$. For any collection $\mathbb{C}$ of clauses in $\mathbb{C}_{A,B}$, let $\Delta(\mathbb{C}) = |\cup_{C\in\mathbb{C}}\Gamma(C)|$. Similarly, define $\Gamma_A(C) = \Gamma(C) \setminus A$ and $\Delta_A(\mathbb{C}) = |\cup_{C\in\mathbb{C}}\Gamma_A(C)|$. We make the following claim:

Claim 3.4.8. For any $\mathbb{C} \subseteq \mathbb{C}_{A,B}$, $\Delta_A(\mathbb{C}) \geqslant (k - 5/2 - 2\beta)|\mathbb{C}|$.

We first complete the proof of the lemma using the claim. Since $\Delta_A(\mathbb{C}_{A,B}) \geqslant (k - 5/2 - 2\beta)|\mathbb{C}_{A,B}|$ and $\beta < 1/10$, there exists a clause C such that $|\Gamma_A(C)| \geqslant k - 2$. Now $V(C) \setminus A \supseteq \Gamma_A(C)$ and thus $|V(C) \setminus A| \geqslant k - 2$. We place this clause at the beginning of the ordering, call it C_1 and set $F_1 = V(C) \setminus A$. We now iterate with $\mathbb{C}_{A,B} \setminus \{C\}$ to complete the construction, obtain a clause $C_2 \in \mathbb{C}_{A,B} \setminus C_1$ such that $|\Gamma_A(C_2)| \geqslant k - 2$. Since $\Gamma_A(C_1)$ cannot intersect $\Gamma_A(C_2)$, we can now set $F_2 = V(C_2) \setminus V(C_1)$. Continuing this way yields the required ordering and partition of $B \setminus A$. □

We now complete the proof of the claim.

Proof of Claim 3.4.8. Fix any $\mathbb{C}$ and consider any (maximally) connected subgraph with edges $\mathbb{C}' \subseteq \mathbb{C}$. If $\mathbb{C}'$ consists of a single clause C, then $|V(C) \cap A| \leqslant 1$ (since A is closed) and $V(C) \cap V(C') = \emptyset$ for any $C' \neq C \in \mathbb{C}$. Thus, $\Gamma_A(\mathbb{C}') \geqslant k - 1$.

Now suppose $\mathbb{C}'$ consists of at least 2 clauses. We first claim that $\Delta(\mathbb{C}') \geqslant (k - 2 - 2\beta)|\mathbb{C}'|$. To see this, observe that $\mathbb{C}'$ is a collection of at most ηn clauses in $\mathcal{I}$ and thus, $|V(\mathbb{C}')| \geqslant (k - 1 - \beta)|\mathbb{C}|$. Further, every $v \in V(\mathbb{C}') \setminus \cup_{C \in \mathbb{C}'} \Gamma(C)$ belongs to at least two different clauses in $\mathbb{C}'$ and thus, $(k - 1 - \beta)|\mathbb{C}'| \leqslant |V(\mathbb{C}')| \leqslant \Delta(\mathbb{C}') + (k|\mathbb{C}'| - \Delta(\mathbb{C}'))/2$. Rearranging gives $\Delta(\mathbb{C}') \geqslant (k - 2 - 2\beta)|\mathbb{C}'|$.

Next, we claim that that for every $v \in V(\mathbb{C}') \cap A$ there exists a pair of clauses C, C' such that $V(C \cup C') \cap A = \{v\}$. Consider any clause $C \in \mathbb{C}$ such that $V(C) \cap A = \{v\}$. If there is another clause C' such that $V(C') \cap A = \{v\}$, then observe that $V(C')$ cannot intersect A in any other element (since A is closed) and thus we can let C, C' be the pair as above, corresponding to v. Otherwise, there exists a clause C' such that $C' \in \mathbb{C}$ such that $V(C') \cap V(C) \neq \emptyset$ (since $V(\mathbb{C}')$ is connected) and $V(C') \cap A = \emptyset$ (as otherwise there would

be a path between two distinct vertices of A, of length at most 2 outside of A). Further, observe that all such pairs are disjoint. This is because if some pairs intersect, then they induce a path of length at most 3 between two distinct vertices of A that is not contained in A (violating the 3 closedness of A). Thus, $|V(\mathbb{C}') \cap A| \leqslant |\mathbb{C}'|/2$. Thus, we must have: $\Delta_A(\mathbb{C}') \geqslant \Delta(\mathbb{C}') - |\mathbb{C}'|/2 \geqslant (k - 2 - 2\beta)|\mathbb{C}'| - |\mathbb{C}'|/2 = (k - 5/2 - 2\beta)|\mathbb{C}'|$.

Since for every connected component $\mathbb{C}'$ inside $\mathbb{C}$ we have that $\Delta_A(\mathbb{C}') \geqslant (k - 5/2 - 2\beta)|\mathbb{C}'|$, we must have $\Delta_A(\mathbb{C}) \geqslant (k - 5/2 - 2\beta)|\mathbb{C}|$ as promised. This completes the proof of claim. □

3.4.3 $\tilde{\mathbb{E}}$ and some basic properties

The following is immediate from (3.4.1):

Lemma 3.4.9. *Suppose A and B are closed disjoint sets such that $A \cup B$ is closed. Then, $\nu_{A\cup B}(x) = \nu_A(x_A) \cdot \nu_B(x_B)$.*

We now define the pseudo-expectation operator associated with the local distributions $\{\nu_T\}_{|T|\leqslant s}$:

Definition 3.4.10 (Pseudo-Expectation)**.** For the collection of consistent local probability distributions $\{\nu_T\}_{|T|\leqslant s}$ defined in (3.4.1) for $s \leqslant \eta n/6$, we define $\tilde{\mathbb{E}}$ on P^n_s by

$$\tilde{\mathbb{E}}[\chi_S] = \mathop{\mathbb{E}}_{\nu_S}[\chi_S],$$

for every $|S| \leqslant s$.

Corollary 3.4.11. *Let $\mathcal{I}$ be a nice (k, n) instance and μ a pairwise independent distribution over $\{\pm 1\}^k$. Then the family of local distributions $\{\nu_X\}_{X\subseteq[n],|X|<d}$ for $s = \eta n/6$ satisfies:*

1. *Completeness: For every clause C of $\mathcal{I}$, $\nu_{V(C)} = \mu$.*

2. *Consistency: for every $S \subseteq T \subseteq [n]$, $|T| \leqslant d$, the marginal of ν_T on to S is ν_S.*

Proof. The completeness property follows from (3.4.1) and $\mathbb{C}(V(C)) = \{C\}$. The consistency property follows from Lemma 3.4.6. □

Finally, since $\tilde{\mathbb{E}}$ corresponds to a valid expectation locally, we obtain that $\tilde{\mathbb{E}}$ induces a positive semidefinite (PSD) inner product on any space of functions of a small number of variables.

Lemma 3.4.12 (Local PSDness)**.** *Let $\tilde{\mathbb{E}}$ be the pseudo-expectation operator defined by the local distributions $\{\nu_S\}_{|S|\leqslant s}$. Let T be a subset of $[n]$ of size at most s. Then for every $f \in V = \mathsf{Span}\{\chi_A \mid A \subseteq T\}$, $\tilde{\mathbb{E}}[f^2] \geqslant 0$.*

3.5 Local Distribution on Unions

In this section we make an important step towards showing the positivity property of our pseudo-distribution by showing that if two sets A and B are sufficiently closed, then the local distribution on $A \cup B$ is only determined by the clauses that are contained in A or in B. In particular, this implies that if A and B are disjoint then the distribution on A is *independent* of the distribution of B. The main result of this section is the following expression for the local distribution on the union of A and B where A is R-closed for a sufficiently large constant R and B is closed.

Lemma 3.5.1 (Local Distribution on Union of Two Closures)**.** *Suppose A is R-closed for*

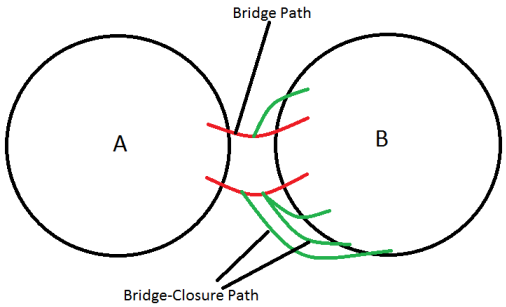


Figure 3.3: A possible configuration when A is R-closed and B is closed. All solid lines indicate paths of length at most 3.

$R \geqslant 100$ and B is closed. Then, for any $x \in \{-1,1\}^{A\cup B}$,

$$\nu_{A\cup B}(x) = Z_{A,B} \cdot \Pi_{C:V(C)\subseteq A\cup B}\mu_C(x_C),$$

where $Z_{A,B} = 2^{k|\mathbb{C}(A\cup B)|-|A\cup B|}$.

We make two convenient definitions before proceeding, see Figure 3.3:

Definition 3.5.2 (Bridge Paths)**.** For any two closed sets A and B, a path P of length at most 3 is said to be a bridge path for the pair A, B if $|P \cap A| = |P \cap B| = 1$.

Definition 3.5.3 (Bridge-Closure Paths)**.** For any two closed sets A and B, a path P of length at most 3 is said to be a bridge-closure path for the pair A, B, if there exists a bridge path P' such that $|P' \cap P| = 1$ and $|P \cap B| = 1$ but $C \cap A = \emptyset$.

Proof overview.. Since the proof is rather technical, let us start with a high level overview of it. We first show the only extra clauses added to $\mathsf{cl}(A \cup B)$ come from bridge and bridge-closure paths. Moreover, all these additional paths are disjoint apart from their end points.

What this amounts to is that the new connections between A and B can be thought of as a collection of disjoint *trees* $T_1, \ldots, T_r$ such that each of these trees has a root in A and its leaves in B. The marginal distribution over $A \cup B$ is obtained by summing up all possible assignments to the intermediate nodes in these trees. Thus at the heart of the proof is the observation that for every such tree T with root x_0 and leaves $x_1, \ldots, x_\ell$, if we consider the distribution over the variables of T induced by the tree (i.e., where the probability of x is proportional to $\prod_{C \in \mathbb{C}(T)} \mu_C(x_C)$) then the marginal distribution over $\{x_0, x_1, \ldots, x_\ell\}$ is uniform. Hence these trees create no dependence between A and B.

As a final remark, observe that the example from Figure 3.1 shows that A and B being 2-closed is not enough to guarantee the statement of the lemma. While we believe that at least one of the sets out of A and B should be R-closed for some $R > 3$ for the lemma to hold, currently, we do not have any example of a counter example demonstrating this point. We now proceed with the actual proof.

Proof of Lemma 3.5.1. Let $D = \mathsf{cl}(A \cup B)$. Let $\mathbb{C}_{A,B}$ and $\mathbb{C}_B$ be the set of bridge paths and bridge closure paths of B for the pair A, B, respectively. Observe that $V(\mathbb{C}_{A,B}) \cup V(\mathbb{C}_B) \subseteq D$. We now show that these are the only *extra* clauses in D:

We first make a few simple observations:

The first observation describes how bridge paths and bridge-closure paths intersect.

Claim 3.5.4 (Intersections).

1. For any distinct $P_1, P_2 \in \mathbb{C}_{A,B}$, $P_1 \cap P_2 \subseteq A \cup B$.
2. For any distinct $P_1, P_2 \in \mathbb{C}_B$, $P_1 \cap P_2 \subseteq V(P) \cup B$ where P is a bridge path.
3. For any $P \in \mathbb{C}_B$ and $P' \in \mathbb{C}_{A,B}$, $|V(P) \cap V(P')| \leqslant 1$.

4. Suppose $P_1, P_2 \in \mathbb{C}_B$ are such that $P_1 \cap P \neq \emptyset$ and $P_2 \cap P' \neq \emptyset$ for some bridge paths $P \neq P'$. Then, $P_1 \cap P_2 = \emptyset$.

Proof. 1. If the claim weren't true, then there must be a path of length $\leqslant 6$ between two vertices of A which violating that A is R-closed.

2. Suppose first that there is a bridge path P such that $P \cap P_1 \neq \emptyset$ and $P \cap P_2 \neq \emptyset$. If either of P_1 or P_2 intersect P in more than one element, then there is a cycle of length at most 6 in G which violates the fact that G has $\Omega(1)$ girth. If P_1 and P_2 intersect in an element not contained in $V(P)$, then, again there is a cycle of length at most 9 in G violating the high girth of G. Similarly, if P_1, P_2 intersect inside B, then, they cannot intersect outside of B and further, they cannot both intersect the same bridge path as it would yield a cycle of length at most 9 in G. Thus in both the cases, $P_1 \cap P_2 \subseteq V(P) \cup B$ for some bridge path P.

3. Otherwise there is a cycle of length at most 6 in G violating that G has girth $\omega(1)$.

4. If not, then if $|P \cap P' \cap A| = 1$ then there is a cycle of length 12 in the graph, contradicting our assumption on the girth. Otherwise $|P \cap P' \cap A| = 2$ which means that there is a path of length at most 12 between two distinct vertices of A.

□

The next observation shows that there is no path of length at most 3 that connects two bridge paths, two bridge-closure paths or two bridge-bridge-closure paths that are not contained in $A \cup B$.

Claim 3.5.5 (No Extra Paths). 1. There is no path P of length at most 3 not contained in A that connects a bridge path P' and A.

2. There is no path of length at most 3 not contained in A that connects $P \in \mathbb{C}_{A,B}$ with $P' \in \mathbb{C}_B$.

3. There is no path of length at most 3 connecting distinct $P, P' \in \mathbb{C}_B$.

Proof. 1. Otherwise there is a path of length at most 6 between two vertices of A not contained in A, violating the fact that A is R closed.

2. Otherwise there is a path of length at most 12 between two vertices of A, violating that A is R closed.

3. Otherwise there is a path of length at most 18 not contained in A, connecting two vertices of A.

□

The following claim is now a consequence of the claims above:

Claim 3.5.6. For any C such that $V(C) \not\subseteq A \cup B$ but $V(C) \subseteq D$, $C \in \mathbb{C}_{A,B} \cup \mathbb{C}_B$.

Proof of Claim. Consider the iterative procedure of building the closure of $A \cup B$ by adding paths one by one in some order. Let P be the first path in this order that violates the claim. Then, either P intersects two bridge paths or a bridge path and A or a bridge path and a bridge-closure path or two bridge-closure paths. Each of these situations is explicitly barred by the claims above. This completes the argument. □

Let $Z = 2^{k|\mathbb{C}(D)|-|D|} = 2^{k|\mathbb{C}(A\cup B)|+k|\mathbb{C}_{A,B}|-|D|}$. Observe that $Z \cdot 2^{-2|\mathbb{C}_{A,B}|} = Z_{A,B}$. For every clause $C \in \mathbb{C}_{A,B} \cup \mathbb{C}_B$, let $V'_C = V(C) \setminus (A \cup B)$ and $V''_C = V(C) \cap (A \cap B)$. Similarly, let $D' = D \setminus (A \cup B)$ and $D'' = D \cap (A \cup B)$. Next, we claim:

Claim 3.5.7.

$$Z \cdot \sum_{\gamma \in \{-1,1\}^{D'}} \Pi_{C \in \mathbb{C}_{A,B} \cup \mathbb{C}_B} \mu_C \left(x_{V''_C} \circ \gamma_{V'_C} \right) = Z_{A,B}.$$

Proof. Let $D' = V_1 \cup V_2$ such that $V_1 \cap V_2 = \emptyset$ defined by $V_1 = D' \setminus V(\mathbb{C}_{A,B}))$ and $V_2 = D' \setminus V_1$.

$$\begin{aligned}
&Z \cdot \sum_{\gamma \in \{-1,1\}^{D'}} \Pi_{C \in \mathbb{C}_{A,B} \cup \mathbb{C}_B} \mu_C \left(x_{V''_C} \circ \gamma_{V'_C} \right) \\
&= Z \sum_{\gamma \in \{-1,1\}^{D'}} \Pi_{C \in \mathbb{C}_B} \mu_C \left(x_{V''_C} \circ \gamma_{V'_C} \right) \Pi_{C \in \mathbb{C}_{A,B}} \mu_C \left(x_{V''_C} \circ \gamma_{V'_C} \right) \\
&= Z \sum_{\gamma \in \{-1,1\}^{V_2}} \Pi_{C \in \mathbb{C}_{A,B}} \mu_C \left(x_{V''_C} \circ \gamma_{V'_C} \right) \sum_{\gamma \in \{-1,1\}^{V_1}} \Pi_{C \in \mathbb{C}_B} \mu_C \left(x_{V''_C} \circ \gamma_{V'_C} \right)
\end{aligned}$$

Now, observe that for every $C \in \mathbb{C}_B$, $V(C) \cap V_2$ has at most 2 elements. Thus, by pairwise independe

$$= Z \sum_{\gamma \in \{-1,1\}^{V_2}} \Pi_{C \in \mathbb{C}_{A,B}} \mu_C \left(x_{V''_C} \circ \gamma_{V'_C} \right) \Pi_{C \in \mathbb{C}_B} 2^{-|V(C) \cap (A \cup B \cup V_1)|}$$

Similarly, for every $C \in \mathbb{C}_{A,B}$, $V(C) \cap (A \cup B)$ contains at most 2 elements. Thus,

$$= Z \Pi_{C \in \mathbb{C}_{A,B}} 2^{-|V(C) \cap (A \cup B)|} \Pi_{C \in \mathbb{C}_B} 2^{-|V(C) \cap (A \cup B \cup V_1)|}$$

$$= Z_{A,B}.$$

□

We can now write, using (3.4.1):

$$\nu_{A\cup B}(x) = Z \cdot \sum_{\gamma\in\{-1,1\}^{D'}} \Pi_{C:V(C)\subseteq D}\mu_C\left(x_{V_C''}\circ\gamma_{V_C'}\right)$$

Using Claim 3.5.6

$$= Z \sum_{\gamma\in\{-1,1\}^{D'}} \Pi_{C:V(C)\subseteq(A\cup B)}\mu_C(x_C)\cdot\Pi_{C\in\mathbb{C}_{A,B}\cup\mathbb{C}_B}\mu_C\left(x_{V_C''}\circ\gamma_{V_C'}\right)$$

Using Claim 3.5.7

$$= Z_{A,B}\Pi_{C:V(C)\subseteq(A\cup B)}\mu_C(x_C).$$

This completes the proof. □

3.6 $\tilde{\mathbb{E}}$ is positive semidefinite

In this section, we prove our main result. Our proof will follow easily from the following lemma which is the main result of this section.

Lemma 3.6.1 (Main Lemma). *Let $P_d^n = \mathsf{Span}\{\chi_A \mid |A| \leqslant d\}$ be the space of multilinear polynomials on $\mathbb{R}^n$ of degree at most $d = \frac{\eta n}{10000k}$. There exists a collection of functions $\{\tilde{\chi}_i \mid 0 \leqslant i \leqslant M\} \subseteq P_d^n$ for $M = \binom{n}{\leqslant d} - 1$ such that:*

1. $P_d^n = \mathsf{Span}\{\tilde{\chi}_i \mid 0 \leqslant i \leqslant M\}$.
2. $\tilde{\mathbb{E}}[\tilde{\chi}_i^2] \geqslant 0$.
3. $\tilde{\mathbb{E}}[\tilde{\chi}_i \cdot \tilde{\chi}_j] = 0$ *whenever* $i \neq j$.

We first complete the proof of of Theorem 3.0.2 assuming this lemma. Observe that part (1) of the theorem follows from Lemma B.2.3. Further, $\tilde{\mathbb{E}}$ satisfies $\tilde{\mathbb{E}}[f(C_i)] = f(\mu)$

by Corollary 3.4.11. Thus, we only need to prove that $\tilde{\mathbb{E}}$ is a valid pseudo-expectation operator, that is, that $\tilde{\mathbb{E}}$ is positive semidefinite.

Let $f \in P_d^n$ be any multilinear polynomial of degree $\leqslant d$. Then, we show that $\tilde{\mathbb{E}}[f^2] \geqslant 0$. We use the spanning property (1) of the $\tilde{\chi}_i$s above to write $f = \sum_{i<\binom{n}{\leqslant d}} f_i \cdot \tilde{\chi}_i$. Using orthogonality (3) of $\tilde{\chi}_i$s, we have: $\tilde{\mathbb{E}}[f^2] = \sum_{i\in\binom{[n]}{\leqslant d}} f_i^2 \tilde{\mathbb{E}}[\tilde{\chi}_i^2]$. Finally, using the positivity property (2) of the $\tilde{\chi}_i$s, we have that $\tilde{\mathbb{E}}$ is PSD.

The rest of this section is devoted to proving Lemma 3.6.1.

3.6.1 Choosing an Ordering

Our aim is to build an order on the $\binom{[n]}{\leqslant d}$, in which to process them for our local orthogonalization procedure. We start with an arbitrary ordering on the clauses of $\mathcal{I}$, e.g. for every $C \in \mathcal{I}$ we define a unique index $\zeta(C) \in [m]$. We say that $A \prec B$ if:

- $\mathbb{C}(\mathsf{cl}(A))$ is smaller than $\mathbb{C}(\mathsf{cl}(B))$ in lexicographic order of ζ. That is, $A \prec B$ if the maximum index $\zeta(C)$ for $C \in \mathsf{cl}(A)$ is smaller than this maximum for $\mathsf{cl}(B)$, and if they are equal we break ties by the second largest index and so on. We define $\pi(\mathsf{cl}(A))$ to be the index of $\mathsf{cl}(A)$ according to this ordering. (Note that π is a permutation on *distinct* closures, and so if $\mathsf{cl}(A) \neq \mathsf{cl}(B)$ then $\pi(\mathsf{cl}(A)) \neq \pi(\mathsf{cl}(B))$.)
- If $\mathbb{C}(\mathsf{cl}(A)) = \mathbb{C}(\mathsf{cl}(B))$ then we say that $A \prec B$ if $|A| < |B|$.
- If $\mathbb{C}(\mathsf{cl}(A)) = \mathbb{C}(\mathsf{cl}(B))$ and $|A| = |B|$ then we break ties arbitrarily.

For $i = 0, \ldots, M$, we let A_i denote the i^{th} set in this ordering. Note that $A_0 = \emptyset$ and $A_1, \ldots, A_n$ are the singleton elements $\{1\}, \ldots, \{n\}$ (in some arbitrary order). We will

write χ_i for χ_{A_i} in the following to reduce clutter.

3.6.2 Local Orthogonalization

Set $R = 100$. Define the i^{th} *local correlated space* as

$$V_i = \mathsf{Span}\{\chi_B \mid |B| \leqslant d, B \subseteq \mathsf{cl}_R(A_i), B \prec A_i\}.$$

Lemma 3.6.2. *For every $f \in V_i$, $\tilde{\mathbb{E}}[f^2] \geqslant 0$.*

Proof. Invoking Lemma 3.4.12, it suffices to show that $|\mathsf{cl}_R(A_i)| < s = \eta n/6$. This follows by noting that $|A_i| \leqslant d$ and $|\mathsf{cl}_R(A_i)| \leqslant 2Rkd = 200\eta n/10000 = \eta n/500 \leqslant s$ (Lemma 3.4.4). □

Define $\overline{\chi}_i$ to be any $f \in V_i$ such that $\tilde{\mathbb{E}}[(\chi_i - f)^2] \leqslant \tilde{\mathbb{E}}[(\chi_i - g)^2]$ for every $g \in V_i$. Note that such a function must exist because $\tilde{\mathbb{E}}[(\chi_i - f)^2] \geqslant 0$ for every f (one can WLOG minimize on the orthogonal complement of the kernel of $\tilde{\mathbb{E}}$ inside V_i). We define $\tilde{\chi}_i = \chi_i - \overline{\chi}_i$. Since V_0 is empty, we set $\overline{\chi}_0$ as the constant 0 function and $\tilde{\chi}_0$ is thus defined as $\chi_0 = \chi_\emptyset = 1$.

The following simple lemma would be very useful.

Lemma 3.6.3 (Local orthogonality)**.** *$\tilde{\mathbb{E}}[\tilde{\chi}_i g] = 0$ for every $g \in V_i$.*

Proof. Since both g and $\overline{\chi}_i$ are spanned by characters of size at most d and $2d < s$, the pseudo-expectation is well defined. Further, since both g and $\overline{\chi}_i$ lie in $\mathsf{Span}\{\chi_S \mid S \subseteq \mathsf{cl}_R(A_i)\}$ and $|\mathsf{cl}_R(A_i)| \leqslant s$ (as in the proof of Lemma 3.6.2), $\tilde{\mathbb{E}}$ corresponds to the expectation operator associated with the probability distribution $\nu_{\mathsf{cl}_R(A_i)}$.

Now suppose for the sake of contradiction that

$$\tilde{\mathbb{E}}[(\tilde{\chi}_i - \overline{\chi}_i)g] = \delta$$

for some $\delta > 0$. (If the expectation is negative then we can take $-g$.) Let $f = \overline{\chi}_i - \varepsilon g$. We have:

$$\tilde{\mathbb{E}}[(\chi_i - f)^2] = \tilde{\mathbb{E}}[(\chi_i - \overline{\chi}_i)^2] + \varepsilon^2\tilde{\mathbb{E}}[g^2] - 2\varepsilon\delta$$

and so if ε is sufficiently small then

$$\tilde{\mathbb{E}}[(\chi_i - f)^2] < \tilde{\mathbb{E}}[(\chi_i - \overline{\chi}_i)^2]$$

contradicting our choice of $\overline{\chi}_i$. □

The following lemma shows that the $\tilde{\chi}_i$'s span P_d^n:

Lemma 3.6.4. *For every i:* $\mathsf{Span}\{\tilde{\chi}_j : j \leqslant i\} = \mathsf{Span}\{\chi_j : j \leqslant i\}$.

Proof. First, we show that for every i, $\chi_i \in \mathsf{Span}\{\tilde{\chi}_j \mid j \leqslant i\}$. We argue by induction. $\tilde{\chi}_1 = \chi_1$ and thus the statement holds for $i = 1$. Now suppose the statement holds for all $j < i$. Consider χ_i. From the definition of $\tilde{\chi}_i$ above, we have that: $\chi_i = \tilde{\chi}_i + \overline{\chi}_i$. Now, $\overline{\chi}_i \in V_i$ and $V_i \subseteq \mathsf{Span}\{\chi_j \mid j < i\}$ by definition. Further, by inductive hypothesis, each χ_j for $j < i$ satisfies $\chi_j \in \mathsf{Span}\{\tilde{\chi}_{j'} \mid j' \leqslant j\} \subseteq \mathsf{Span}\{\tilde{\chi}_{j'} \mid j' < i\}$. This completes the induction.

The other direction is easier: $\tilde{\chi}_i = \chi_i - \overline{\chi}_i$ and as argued above, $\overline{\chi}_i \in \mathsf{Span}\{\chi_j \mid j < i\}$. Thus, $\tilde{\chi}_i \in \mathsf{Span}\{\chi_j \mid j \leqslant i\}$. Together, we thus have: $\mathsf{Span}\{\tilde{\chi}_j \mid j \leqslant i\} = \mathsf{Span}\{\chi_j \mid j \leqslant i\}$. □

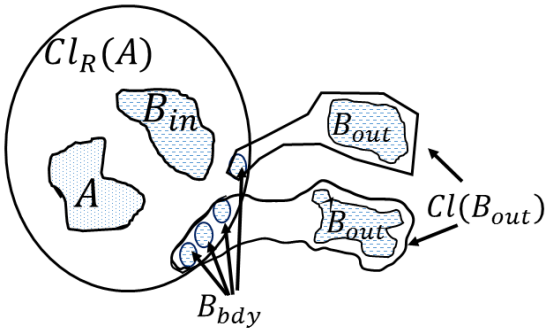


Figure 3.4: A possible configuration of B_{in}, B_{out} and B_{bdy}.

3.6.3 Global Orthogonality lemma

In this section, we prove the following lemma that is the technical heart of the proof and says that local orthogonalization is enough to ensure that $\tilde{\chi}_i$ are all mutually orthogonal.

Lemma 3.6.5. *For every $j < i$,*

$$\tilde{\mathbb{E}}[\tilde{\chi}_i \cdot \chi_j] = 0.$$

We will need the following observation for the proof which we record before proceeding:

Lemma 3.6.6. *Suppose H is a connected k-uniform hypergraph such that there exist a subset of vertices, U, $|U| \geqslant 2$ satisfying:* $\mathrm{dist}(u, v) > R$ *for every distinct $u, v \in U$. Then, H must have at least $\frac{|U|R}{2}$ hyperedges.*

Proof. Observe that the collection of balls of radius $R/2$ around any vertex in $u \in U$ are all disjoint and contain at least one path (due to connectedness of H). □

We now go on to prove Lemma 3.6.5.

Proof. Fix any $j < i$ and let $A = A_i$ and $B = A_j$. Let

$$B_{bdy} = \{x \in \mathsf{cl}(B) \mid \exists \text{ a clause } v(W) \in \mathbb{C}(\mathsf{cl}(B)) \text{ s.t. } W \cap \mathsf{cl}_R(A) = \{x\}\}.$$

For every $x \in B_{bdy}$ we call any associated clause W as in the definition above as a *boundary clause*. Let $B_{out} = B \setminus \mathsf{cl}_R(A_i)$ and $B_{in} = B \setminus (B_{out} \cup B_{bdy})$ and $B_{rest} = B \setminus (B_{out} \cup B_{in})$. Note that B_{bdy} is not necessarily a subset of B. Next, we make two useful claims:

Claim 3.6.7.

$$\left|B_{bdy} \cup B_{in}\right| \leqslant |B| \leqslant d.$$

Proof. We will show that $|B_{bdy}| \leqslant |B_{out}|$. This immediately yields the claim by observing that $d \geqslant |B| = |B_{in}| + |B_{out}| + |B_{rest}| \geqslant |B_{in}| + |B_{bdy}|$. We note that the proof of this claim is significantly simpler in the case that $|B| < R/2$. Proving it in the case when R is a constant and $|B| = \Omega(n)$ is one of the main technical ingredients in getting the proof sketched in the overview to work for $\Omega(n)$ rounds of the SOS hierarchy.

Let $Q \subseteq [n]$ be a (maximally) connected component in the subgraph defined by the hyperedges $\mathbb{C}(\mathsf{cl}(B)) \setminus \mathbb{C}(\mathsf{cl}_R(A))$. Let $Q_{bdy} = B_{bdy} \cap Q$ and $Q_{out} = B_{out} \cap Q$. B_{bdy} is thus partitioned into Q_{bdy} for every possible maximally connected subgraphs Q. It is thus enough to prove that $|Q_{bdy}| \leqslant |Q_{out}|$ for any fixed Q.

Observe that $Q \cap \mathsf{cl}_R(A) = Q_{bdy}$. If $Q \cap \mathsf{cl}_R(A) = \emptyset$, then, there is nothing to prove. If $Q_{bdy} = \{v\}$, then, Q contains $V(W_v)$ where W_v is a boundary clause associated with v. If

Q contains no vertex of B_{out}, then, observe that $\mathsf{cl}(B) \setminus (Q \setminus \{v\})$ is a closed set containing B contradicting the minimality of $\mathsf{cl}(B)$. Thus, in this case, $|Q_{bdy}| \leqslant |Q_{out}|$.

Now suppose for $|Q_{bdy}| \geqslant 2$. Then, vertices in Q_{bdy} are connected through clauses in Q. On the other hand, since A is R-closed, for any $u, v \in Q_{bdy}$, any path that uses clauses from Q between u, v must be of length at least $R+1$. Applying Lemma 3.6.6, we observe that $|\mathbb{C}(Q)| \geqslant |Q_{bdy}|R/2$.

Next, we claim that $Q \subseteq \mathsf{cl}(Q_{bdy} \cup Q_{out})$. It is easy to complete the proof once we have this claim: observe that

$$|Q_{bdy}|R/2 \leqslant |\mathbb{C}(Q)| \leqslant |\mathbb{C}(\mathsf{cl}(Q_{bdy} \cup Q_{out}))| \leqslant 6|Q_{bdy}| + 6|Q_{out}|.$$

Rearranging yields that $|Q_{out}| \geqslant |Q_{bdy}| \cdot \frac{R/2-6}{6}$. Using $R \geqslant 24$ yields that $|Q_{out}| \geqslant |Q_{bdy}|$.

We now proceed to show that $Q \subseteq \mathsf{cl}(Q_{bdy} \cup Q_{out})$. By Lemma 3.4.5 (4), $\mathsf{cl}(B) = \mathsf{cl}(B_{in} \cup B_{bdy} \cup B_{out})$. Let $B' = B_{in} \cup B_{bdy} \cup B_{out} \setminus (Q_{bdy} \cup Q_{out})$. Then, by another application of Lemma 3.4.5 (4), $\mathsf{cl}(B) = \mathsf{cl}(\mathsf{cl}(B') \cup \mathsf{cl}(Q_{bdy} \cup Q_{out}))$. In other words, one can build the closure of B by first building the closure of B' and $Q_{bdy} \cup Q_{out}$ (Step 1) and then taking the closure of the unions of the obtained sets (Step 2). Clearly, the final output contains every clause in $\mathbb{C}(Q)$. If we show that (1) $\mathbb{C}(\mathsf{cl}(B')) \cap \mathbb{C}(Q) = \emptyset$ and that (2) no clause from $\mathbb{C}(Q)$ is added in the step 2, then every clause in $\mathbb{C}(Q)$ must be added in the procedure to build $\mathsf{cl}(Q_{bdy} \cup Q_{out})$ and thus we are done. We now proceed to show the two statements above.

(1): First observe that $\mathsf{cl}(B')$ itself can be built by building the closure of B_{in} (and $\mathsf{cl}(B_{in}) \subseteq \mathsf{cl}_R(A) \Rightarrow \mathbb{C}(\mathsf{cl}(B_{in})) \cap \mathbb{C}(Q) = \emptyset$), the closure of $B_{out} \cup B_{bdy} \setminus (Q_{bdy} \cup Q_{out})$ (that cannot intersect any clause from $\mathbb{C}(Q)$ as then Q must include a vertex from $B_{out} \cup B_{bdy} \setminus$

$(Q_{bdy} \cup Q_{out})$, a contradiction) and finally taking the closure of their union. This last step cannot add a clause in Q: every path P added connects $\mathsf{cl}(B_{in})$ and $\mathsf{cl}(B_{out} \cup B_{bdy} \setminus (Q_{bdy} \cup Q_{out}))$. If P is contained in $\mathsf{cl}_R(A)$, then, there is nothing to prove. Otherwise P must pass (exactly once) through a boundary vertex. If P contains a clause from $\mathbb{C}(Q)$, then, if P passes through a boundary vertex not in Q_{bdy}, then this enlarges Q violating that Q is a maximally connected component. If, on the other hand, P passes through a boundary vertex in Q_{bdy}, then, P connects $B_{out} \setminus Q$ with Q violating the maximality of Q. Thus, $\mathbb{C}(\mathsf{cl}(B'))$ cannot include any clause from $\mathbb{C}(Q)$.

(2): Consider the step 2 of the procedure to build $\mathsf{cl}(B)$. In this step, we add paths (of length at most 3) that connect $\mathsf{cl}(B')$ and $\mathsf{cl}(Q_{bdy} \cup Q_{out})$. For any such path P, if P includes some clause C from $\mathbb{C}(Q)$ then it crosses out of $\mathsf{cl}_R(A)$ (exactly once) and thus must pass through a boundary vertex. By maximality of Q, we must have that $P \cap B_{bdy} \in Q_{bdy}$ and $P \setminus \mathbb{C}(\mathsf{cl}_R(A)) \subseteq \mathbb{C}(Q)$. On the other hand, the part of P that connects some vertex in Q_{bdy} to $\mathsf{cl}(Q_{bdy} \cup Q_{out})$ is of length at most 3 and thus must be contained in $\mathsf{cl}(Q_{bdy} \cup Q_{out})$. Thus every edge in $P \setminus \mathbb{C}(\mathsf{cl}_R(A))$ is present in $\mathbb{C}(\mathsf{cl}(Q_{bdy} \cup Q_{out})$ and thus $C \in \mathbb{C}(Q)$.

□

Claim 3.6.8. Suppose $B_{out} \neq \emptyset$. Then, for every $S \subseteq B_{in} \cup B_{bdy}$, $S \prec A$.

Proof. Since $B_{out} \neq \emptyset$, $\mathsf{cl}(B) \neq \mathsf{cl}(A)$. Thus, $\pi(\mathsf{cl}(B)) < \pi(\mathsf{cl}(A))$. Now, $|B_{in} \cup B_{bdy}| \leqslant d$ from Claim 3.6.7. Thus, every subset $S \subseteq B_{in} \cup B_{bdy}$ has a well-defined ordering w.r.t $\binom{[n]}{\leqslant d}$. Further, for every such S, $\mathsf{cl}(S) \subseteq \mathsf{cl}(B)$ (Lemma 3.4.5) and thus, $\pi(\mathsf{cl}(S)) \leqslant \pi(\mathsf{cl}(B))$. Hence, $S \prec A$. □

We now proceed to complete the proof of the lemma. It is easy to verify that $|\mathsf{cl}_R(A) \cup \mathsf{cl}(B)| < s$ and thus by Corollary 3.4.11 the $\tilde{\mathbb{E}}$ operator on functions on variables in $\mathsf{cl}_R(A) \cup \mathsf{cl}(B)$ corresponds to the expectation of a valid local distribution. In what follows, whenever we write $\mathbb{P}$, we mean the probability of an event w.r.t. this local probability distribution. (Note that the expectation w.r.t. this probability distribution agrees with $\tilde{\mathbb{E}}$ whenever both are defined.)

Now, $\chi_B = \chi_{B_{in}} \chi_{B_{rest}} \chi_{B_{out}}$ and we can write

$$\tilde{\mathbb{E}}[\tilde{\chi}_i \chi_j] = \tilde{\mathbb{E}}[\tilde{\chi}_i \chi_{B_{in}} \chi_{B_{rest}} \chi_{B_{out}}] \tag{3.6.1}$$

Consider an arbitrary assignment z to $B \setminus A$ and $\gamma \in \{\pm 1\}^{|B_{bdy}|}$ to $x_{B_{bdy}}$. Let $\mathbb{1}_{B_{bdy}=\gamma}$ be the function that on input $x \in \{\pm 1\}^n$ outputs 1 if $x_{B_{bdy}} = \gamma$ and zero otherwise.

Lemma 3.5.1 gives the expression for the local distribution on $\mathsf{cl}_R(A) \cup \mathsf{cl}(B)$. Using the expression, we have:

$$\tilde{\mathbb{E}}[\tilde{\chi}_i \chi_{B_{in}} \chi_{B_{rest}} \chi_{B_{out}} \mid x_{B \setminus A} = z, x_{B_{bdy}} = \gamma] = \tilde{\mathbb{E}}[\tilde{\chi}_i \chi_{B_{in}} \chi_{B_{rest}} \mid x_{B_{bdy}} = \gamma] \cdot \chi_{B_{out}}(z_{B_{out}}),$$

where the $\tilde{\mathbb{E}}$ on the RHS matches the expectation operator associated with the probability distribution $\nu_{\mathsf{cl}_R(A)}$.

We will show that $\tilde{\mathbb{E}}[\tilde{\chi}_i \chi_{B_{in}} \chi_{B_{rest}} \mid x_{B_{bdy}} = \gamma] = 0$ for every choice of γ. First, we observe that:

$$\mathbb{P}[\mathbb{1}_{B_{bdy}=\gamma}] \cdot \tilde{\mathbb{E}}[\chi_S \cdot \chi_{B_{in}} \chi_{B_{rest}} \mid \mathbb{1}_{B_{bdy}=\gamma}] = \tilde{\mathbb{E}}[\chi_S \cdot \chi_{B_{in}} \chi_{B_{rest}} \cdot \mathbb{1}_{B_{bdy}=\gamma}], \tag{3.6.2}$$

for ever $S \subseteq \mathsf{cl}_R(A)$, $|S| \leqslant d$.

Now, $\tilde{\chi}_i \in \mathsf{Span}\{\chi_S \mid S \subseteq \mathsf{cl}_R(A_i), |S| \leqslant d\}$, and thus using (3.6.2),

$$\mathbb{P}[\mathbb{1}_{B_{bdy}=\gamma}] \cdot \tilde{\mathbb{E}}[\tilde{\chi}_i \chi_{B_{in}} \chi_{B_{rest}} \mid x_{B_{bdy}} = \gamma] = \tilde{\mathbb{E}}[\tilde{\chi}_i \cdot \chi_{B_{in}} \chi_{B_{rest}} \cdot \mathbb{1}_{B_{bdy}=\gamma}].$$

Now $|B_{in} \cup B_{bdy}| \leqslant |B_{out}|$ (Claim 3.6.7) and $\mathbb{1}_{B_{bdy}} = \gamma \in \mathsf{Span}\{\chi_S \mid S \subseteq B_{bdy}\}$:

$$\chi_{B_{in}} \chi_{B_{rest}} \cdot \mathbb{1}_{B_{bdy}=\gamma} \in \mathsf{Span}\{\chi_{B_{in}} \cdot \chi_{B_{rest}} \cdot \chi_T \mid T \subseteq B_{bdy}\}.$$

Each index set S of the characters above is a subset of B and thus $S \prec A_i$ (invoking Claim 3.6.8 along with the fact $\pi(\mathsf{cl}(B)) < \pi(\mathsf{cl}(A))$). Thus, $\chi_{B_{in}} \chi_{B_{rest}} \cdot \mathbb{1}_{B_{bdy}=\gamma} \in V_i$. Using Lemma 3.6.3, thus,

$$\tilde{\mathbb{E}}[\tilde{\chi}_i \cdot \chi_{B_{in}} \chi_{B_{rest}} \cdot \mathbb{1}_{B_{bdy}=\gamma}] = 0.$$

□

We can now complete the proof of Lemma 3.6.1.

Proof of Lemma 3.6.1. We show that the $\tilde{\chi}_i$ constructed above satisfy all the properties required. By Lemma 3.6.4, $\mathsf{Span}\{\tilde{\chi}_i \mid i \leqslant M\} = \mathsf{Span}\{\chi_i \mid i \leqslant M\} = P^n_d$. Next, observe that $\tilde{\chi}_i = \chi_i - \overline{\chi}_i$. Both χ_i and $\tilde{\chi}_i$ lie in $\mathsf{Span}\{\chi_S \mid S \subseteq \mathsf{cl}_R(A_i)\}$ and by Lemma 3.4.12, $\tilde{\mathbb{E}}$ is a psd expectation operator over V_i. Thus, $\tilde{\mathbb{E}}[\tilde{\chi}_i^2] \geqslant 0$ for every $i \leqslant M$. Finally, we verify that $\tilde{\chi}_i$ are mutually orthogonal. Fix any i. It is then enough to show that $\tilde{\mathbb{E}}[\tilde{\chi}_j \cdot \tilde{\chi}_i] = 0$ for every $j \neq i$. Since $\mathsf{Span}\{\tilde{\chi}_r \mid r \leqslant j\} = \mathsf{Span}\{\chi_r \mid r \leqslant j\}$ (Lemma 3.6.4), we only need to show that $\tilde{\mathbb{E}}[\chi_j \cdot \tilde{\chi}_i] = 0$ for every $j < i$. Invoking Lemma 3.6.5 then completes the proof. □

Chapter 4

Optimal Sum of Squares Lower Bound for Planted Clique at Degree Four and Tight Analysis of Simple Moments

In this chapter[1], we prove a tight lower bound on the sum of squares algorithm of degree four for finding planted cliques in random graphs. We also give a tight analysis of the construction of [85]. The results of this chapter were obtained in joint work with Samuel Hopkins and Aaron Potechin [66].

The main result of this chapter is the following lower bound on the degree four sum of squares algorithm for the planted clique problem:

Theorem 4.0.1 (Degree 4 Lower Bound)**.** *The canonical degree* 4 *SoS relaxation of the planted clique problem ((4.0.1)) has an integrality gap of at least* $\tilde{O}(\sqrt{n})$ *with high probability.*[2]*.*

We also give a tight analysis of the certificate considered by [85] and [43] and show that it yields a lower bound of $n^{\frac{1}{\frac{d}{2}+1}}$ for degree d SoS relaxation.

[1]The results of this chapter were published in the Proceedings of Symposium on Discrete Algorithms, SODA, 2016 in a paper titled *On the Integrality Gap of Degree-4 Sum of Squares for Planted Clique* co-authored with Samuel Hopkins, Aaron Potechin, Prasad Raghavendra and Tselil Schramm. The conference version is a merge of two papers: *SoS and Planted Clique: Tight Analysis of MPW Moments at all Degrees and an Optimal Lower Bound at Degree Four* co-authored with Samuel Hopkins and Aaron Potechin (in which all the co-authors contributed equally in producing all the results and are listed in alphabetical order) and *Tight Lower Bounds for Planted Clique in the Degree-4 SOS Program* authored by Prasad Raghavendra and Tselil Schramm.

[2]Throughout this thesis, we use $\tilde{O}$ to hide polylogarithmic factors in n

Theorem 4.0.2 (Tight Analysis of MPW)**.** *For every $d = o(\sqrt{\log(n)})$, the canonical degree d SoS relaxation for the planted clique problem ((4.0.1)) has an integrality gap of at least $\tilde{O}(n^{\frac{1}{\frac{d}{2}+1}})$.*

Curiously, the certificate used in [85, 43] is sufficient to show an $\Omega(\sqrt{n/2^d})$ lower bound for the degree d *LS+* hierarchy [46] (which is a weaker SDP that also runs in time $n^{O(d)}$). However, a generalization of an argument of Kelner that we include in this chapter (see Section 4.8) shows that this is *not* the case for the SoS hierarchy, and our analysis of this certificate is tight. Thus, we can conclude that to get stronger lower bounds for higher degree SOS it is necessary and sufficient to utilize more complicated constructions of certificates than those used for weaker hierarchies.

We begin by an overview of the sum of squares method as applied to the planted clique problem and then state our results.

The Sum-of-Squares Method. The SoS semidefinite programming hierarchy yields a convex programming relaxation for the planted clique problem. That is, we derive from the input graph G a convex program $\mathcal{P}_G$ such that if the graph had a clique of size ω then $\mathcal{P}_G$ is feasible. To show that the program *fails* to solve the planted clique problem with parameter ω, we show that with high probability there is a solution (known as a *certificate*) for the program $\mathcal{P}_G$ even when G is a random graph from $G(n, 1/2)$ (which in particular will not have a clique of size $\gg \log n$).

The solution to degree d hierarchies can be thought of as a vector $X \in \mathbb{R}^{n^d}$. For *linear programming* hierarchies this vector needs to satisfy various linear constraints, while for *semi-definite programming* languages it also needs to satisfy constraints of the form $M \succeq 0$

where M is a matrix where each entry is a linear function of X. In previous SoS lower bounds for problems such as random 3XOR/3SAT, Knapsack, and random constraint-satisfaction problems [59, 95, 20], the certificate X was obtained in a fairly natural way, and the bulk of the work was in the analysis. In fact, in all those cases the certificate used in the SoS lower bounds was the same one that was used before for obtaining lower bounds for weaker hierarchies [46]. The same holds for the previous works for the planted clique problem, where the works of [85, 43] used a natural certificate which is a close variant of the certificate used by Feige and Krauthgamer [46] for LS+ lower bounds, and showed that it satisfies the stronger conditions of the SoS program.

It is known that such an approach *cannot* work to obtain a $\approx \sqrt{n}$ lower bound for the SoS program of degree 4 and higher. That is, this natural certificate does *not* satisfy the conditions of the SoS program. Hence to obtain our lower bound for degree 4 SoS we need to consider a more complicated certificate, that can be thought of as making a global "correction" to the simple certificate of [85, 43]. For higher degrees, we have not yet been able to analyze the corresponding complex certificate, but we are able to give a tight analysis of the simple certificate, showing that it certifies an $\omega \approx n^{1/(d+1)}$ lower bound on degree $2d$ SoS relaxation. The key technical difficulty in both our work and prior ones is analyzing the spectrum of random matrices that have *dependent* entries. Deshpande and Montanari [43] achieved such an analysis by a tour-de-force of case analysis specifically tailored for the degree 4 case. However, the complexity of this argument makes it unwieldy to extend to either the case of the more complex certificate or the case of analyzing the simple certificate at higher degrees. Thus, key to our analysis is a more principled representation-theoretic approach, inspired by Grigoriev [59], to analyzing the spectrum

of these kind of matrices. We hope this approach would be of use in further results for both the planted clique and other problems.

We now give an informal overview of the SoS program for planted clique, the [85, 43] certificate, our correction to it, and our analysis. See Section 4.1 and [29] for details.

4.0.1 The SoS program for MAX CLIQUE

Let $G = G([n], E)$ be any graph with the vertex set $[n]$ and edge set E. The following polynomial equations ensure that any assignment $x \in \mathbb{R}^n$ must be the characteristic vector of an ω-sized clique in G:

$$\begin{aligned} x_i^2 &= x_i \text{ for all } i \in [n] \\ x_i \cdot x_j &= 0 \text{ for all } \{i, j\} \notin E \\ \sum_{i=1}^{n} x_i &= \omega \,. \end{aligned} \tag{4.0.1}$$

There is a related formulation (which we refer to as the "optimization version") where the constraint $\sum_{i=1}^n x_i = \omega$ is not present and instead we have an objective function $\sum_{i=1}^n x_i$ to maximize. This latter formulation is used by [43] in their work. It is also the program of focus in the work of Raghavendra and Schramm [93] who independently of us, show an almost optimal lower bound for the planted clique problem for the case of degree 4 SoS relaxation. A point feasible for (4.0.1) is easily seen to be feasible for the optimization version with value ω and hence using the variant (4.0.1) only makes our results stronger. It is unclear, however, whether an explicit constraint of $\sum_{i=1}^n x_i = \omega$ adds more refutation

power to the program. [3]

The degree d SoS hierarchy optimizes over an object called as degree-d *pseudo-expectation* or *pseudo-distribution*. A *degree-d pseudo-expectation operator* for (4.0.1) is a linear operator $\tilde{\mathbb{E}}$ that behaves to some extent as the expectation operator for some distribution over $x \in \mathbb{R}^n$ that satisfies the conditions (4.0.1). For example this operator will satisfy that $\tilde{\mathbb{E}} \sum_{i=1}^n x_i = \omega$, $\tilde{\mathbb{E}} x_i^2 = \tilde{\mathbb{E}} x_i$, etc.. More formally, $\tilde{\mathbb{E}}$ is a linear operator mapping every polynomial P of degree at most d into a number $\tilde{\mathbb{E}}P$ such that $\tilde{\mathbb{E}}1 = 1$, $\tilde{\mathbb{E}}P^2 \geqslant 0$ for every P of degree at most $d/2$, $\tilde{\mathbb{E}}PQ = 0$ for every Q of degree at most $d - \deg P$ and P such that the constraint $\{P = 0\}$ appears in (4.0.1). Note that since the dimension of the set of n-variate polynomials of degree at most d is at most n^d, the operator $\tilde{\mathbb{E}}$ can be described as a vector in $\mathbb{R}^d$. Moreover, the constraints on $\tilde{\mathbb{E}}$ can be captured by a semidefinite program, and this semidefinite program is in fact the SoS program. See Section 4.1, the survey [29] or the lecture notes [18] for more on the SoS hierarchy.

4.0.2 The "Simple Moments"

To show a lower bound of ω, we need to show that for a random graph G, we can find a degree d pseudo-expectation operator that satisfies (4.0.1). Both the papers [85] and [43] utilize essentially the same operator, which we call here the "simple moments". It is arguably the most straightforward way to satisfy the conditions of (4.0.1), and the bulk of the work is then in showing the positivity constraint that $\tilde{\mathbb{E}}P^2 \geqslant 0$ for every P of degree $\leqslant 2$ (in the degree 4 case). [43] shows that this will hold as long as $\omega \ll n^{1/3}$ and

[3] The reason, as we describe when discussing pseudoexpectations, is that adding $\{p = 0\}$ as a constraint ensures that $\tilde{\mathbb{E}}[qp] = 0$ for every $\deg(q) \leqslant d - \deg(p)$ in addition to $\tilde{\mathbb{E}}[p] = 0$.

an argument of Kelner (see Section 4.0.3 below) shows that this is tight and in fact these simple moments fail to satisfy the positivity conditions for $\omega \gg n^{1/3}$.

To define a degree d pseudo-expectation operator $\tilde{\mathbb{E}}$, we need to choose some basis $\{P_1, \ldots, P_N\}$ for the set of polynomials of degree at most d and define $\tilde{\mathbb{E}}P_i$ for every i. The simplest basis is simply the monomial basis. Moreover, since our pseudo-expectation satisfies the constraints $\{x_i^2 = x_i\}$, we can restrict attention to *multilinear* monomials, of the form $x_S = \prod_{i \in S} x_i$ for some $S \subseteq [n]$. Note also that the constraints $x_i x_j = 0$ for $\{i, j\} \notin E$ imply that we must define $\tilde{\mathbb{E}}x_S = 0$ for every S that is not a clique in G. Indeed, the pseudo-distribution $\{x\}$ is supposed to mimic an actual distribution over the characteristic vectors of ω-sized cliques in G, and note that in any such distribution it would hold that $\tilde{\mathbb{E}}x_S = 0$ when S is not a clique.

The simplest form of such a pseudo-distribution is to set

$$\tilde{\mathbb{E}}x_S = \begin{cases} 0 & S \text{ is not a clique} \\ \alpha_{|S|} & \text{otherwise} \end{cases}$$

where $\alpha_{|S|}$ is a constant depending only on the size of S. We can compute the value $\alpha_{|S|}$ by noting that we need to satisfy $\tilde{\mathbb{E}}(\sum_i x_i)^\ell = \sum_{i_1,\ldots,i_\ell} \tilde{\mathbb{E}}x_{i_1} \cdots x_{i_\ell} = \omega^\ell$ for every $\ell = 1, \ldots, d$. Since there would be about $\binom{n}{\ell} 2^{-\binom{\ell}{2}}$ ℓ-sized cliques in the graph G, the value α_ℓ will be $\approx \left(\frac{\omega}{n}\right)^\ell$.[4]

This pseudo-distribution is essentially the same one used by Feige and Krauthgamer [46] for LS+, where they were shown to be valid for the constraints of

[4]One actually needs to make some minor modifications to these moments to ensure they satisfy exactly the constraint $\sum x_i = \omega$ as is done in [85] and in our technical section. However, these corrections have very small magnitudes and so all the observations below apply equally well to the modified moments, and so we ignore this issue in this informal overview.

this problem as long as $\omega < \sqrt{n/2^{d+1}}$. Initially Meka and Wigderson conjectured that a similar bound holds for the SoS program, or in other words, that the $\binom{n}{\leqslant d/2} \times \binom{n}{\leqslant d/2}$ matrix M where $M_{S,T} = \tilde{\mathbb{E}} x_S x_T$ for every $S, T \subseteq [n]$ of size $\leqslant d/2$ is positive semidefinite as long as $\omega \ll \sqrt{n}$. The Meka-Wigderson conjecture would have held if the off-diagonal part of M, which is a random matrix with *dependent* entries, would have a spectral norm comparable with an *independent* random matrix with entries of a similar magnitude. However, this turns out to fail in quite a strong way. An argument due to Jonathan Kelner, described in [18], shows that (for $d = 4$) the matrix M is *not* positive semidefinite as long as $\omega \gg n^{1/3}$. We review this argument below, as it is instructive for our correction.

4.0.3 There is such a thing as too simple

In the simple moments, every 4-clique S gets the same pseudo-expectation. In some sense these moments turn out to be "too random" in that they fail to account for some structure that the graph possesses. Specifically, for $i \in [n]$, consider the linear function $r_i(x) = \sum_j r_{i,j} x_j$ where $r_{i,j}$ equals +1 when $\{i,j\} \in E$, equals −1 when $\{i,j\} \notin E$, and equals 0 when $i = j$. Now, consider the polynomial $P(x) = \sum_{i=1}^n r_i(x)^4$. For every x that is the characteristic vector of an ω-clique in G, $P(x) \geqslant \omega(\omega-1)^4 \geqslant \omega^5/2$; indeed for every i in the clique, $r_i(x)^4$ would equal $(\omega-1)^4$. On the other hand, for every i, let us consider the *expectation* of $\tilde{\mathbb{E}} r_i(x)^4$ taken over the choice of the random graph G. Note that in a random graph the $r_{i,j}$'s are i.i.d. ±1 random variables, and hence

$$\mathbb{E}\,\tilde{\mathbb{E}} r_i(x)^4 = \mathbb{E}\,\tilde{\mathbb{E}}(\sum_j r_{i,j} x_j)^4 = \sum_{j_1,j_2,j_3,j_4 \neq i} \mathbb{E}\, r_{i,j_1} r_{i,j_2} r_{i,j_3} r_{i,j_4} \tilde{\mathbb{E}} x_{j_1} x_{j_2} x_{j_3} x_{j_4} \,. \tag{4.0.2}$$

Let us group the terms on the RHS of (4.0.2) based on the number of distinct j_k's.

There are $O(n^2)$ terms corresponding to two distinct j_k's, each of them is multiplied by $\alpha_2 \approx \left(\frac{\omega}{n}\right)^2$ and so they contribute a total of $C\omega^2$ to the expectation for some constant C. In the terms corresponding to three or four j_k's, there is always one variable $r_{i,j}$ that is not squared, and hence their contribution to the expectation is zero. There are $O(n)$ terms corresponding to a single j_k, each multiplied by $\frac{\omega}{n}$ and so their total contribution is at most ω. We get that in expectation $\tilde{\mathbb{E}} r_i(x)^4 \leqslant C\omega^2$ for some constant C and by Markov this holds with high probability as well (for some different choice of the constant).

The conclusion is that while for every ω-sized clique x, $P(x) \geqslant \omega^5/2$, the simple moments satisfy that $\tilde{\mathbb{E}} P(X) \leqslant Cn\omega^2$. When $\omega \gg n^{1/3}$ this yields a strong discrepancy between the value the simple moments give P and the value that they should have given, had they corresponded to an actual distribution on ω-sized cliques. This discrepancy can be massaged into a degree 2 polynomial Q such that $\tilde{\mathbb{E}} Q^2 < 0$ for the simple moments when $\omega \gg n^{1/3}$, thus showing that in this case these moments do not satisfy the SoS program.

4.0.4 Fixing the simple moments

Our fix for the simple moments is directly motivated by the example above. We want to ensure that the polynomial P will get a pseudo-expectation of $\approx \omega^5$, and that in fact for every i $\tilde{\mathbb{E}} r_i(x)^4$ will be roughly ω^5/n. The idea is to break the symmetry between different equal-sized cliques and give a significantly higher pseudo-expectation to cliques that are somewhat over-represented by these polynomials. Specifically for every set S, define $r_S = \sum_i \prod_{j \in S} r_{i,j}$. Note that r_S is a sum of n entries in $\{\pm 1\}$, and in a random graph it behaves roughly like a normal variable with mean 0 and variance n. Roughly speaking,

the corrected moments will set

$$\tilde{\mathbb{E}}[x_S] = \alpha_{|S|}(1 + r_S\omega/n)$$

for every clique S. Note that when $\omega = \varepsilon\sqrt{n}$, the correction factor would typically be of the form $1 \pm \Theta(\varepsilon)$.[5]

Computing the pseudo-expectation $\tilde{\mathbb{E}}r_i(x)^4$ under the new moments we again get the expression

$$\sum_{j_1,j_2,j_3,j_4} \mathbb{E}\, r_{\{j_1,\ldots,j_4\}}\tilde{\mathbb{E}}x_{\{j_1,\ldots,j_4\}} \ .$$

If we now focus on the contribution of the n^4 terms where the j_k's are all distinct, we see that each such set S yields the term

$$\mathbb{E}\, r_S^2\alpha_4\omega/n \ .$$

Since $\alpha_4 = C\omega^4/n^4$ we get that $\tilde{\mathbb{E}}r_i(x)^4 = C\omega^5/n$ as desired.

4.0.5 Analyzing the corrected moments

The above gives some intuition why the corrected moments might be better than the simple moments for one set of polynomials. But a priori it is not at all clear that those polynomials encapsulated all the issues with the simple moments. Moreover, it is also unclear whether or not the correction itself could introduce additional issues, and create new types of negative eigenvectors. Ruling out these two possibilities is the crux of our analysis.

[5] While it might seem that there is a chance for these pseudo-expectations to be negative, if $\omega < \sqrt{n}/\operatorname{polylog}(n)$ then it is exceedingly unlikely that there will exists an S such that $|r_S| > n/\omega$, and so we ignore this issue in this overview.

Here we discuss some key points from our analysis of $\tilde{\mathbb{E}}$. Since [43] carried out a thorough analysis of the degree-4 simple moments, we begin by reviewing their approach.

Approach of [43]. The PSDness of $\tilde{\mathbb{E}}$ reduces to proving PSDness of a related matrix $\mathcal{M} \in \mathbb{R}^{\binom{[n]}{2} \times \binom{[n]}{2}}$, $\mathcal{M}(S,T) = \tilde{\mathbb{E}}\, x^S x^T$. The eigendecomposition of $E = \mathbb{E}_G[\mathcal{M}]$ has three eigenspaces V_0, V_1, V_2 and eigenvalues $\lambda_0 \approx \omega^4/n^2$ on V_0, $\lambda_1 \approx \omega^3/n^2$ on V_1, and $\lambda_2 \approx \omega^2/n^2$ on V_2. Next, write $\mathcal{M}$ as a block matrix with blocks $\mathcal{M}_{ij} = \Pi_{V_i} \mathcal{M} \Pi_{V_j}$ where Π_V projects to the subspace V. On the diagonal blocks, E contributes large positive eigenvalues. If the on-diagonal blocks $M_{ii} \succeq \lambda_i I$ for some λ_i's so that $\|M_{ij}\| \ll \sqrt{\lambda_i \lambda_j}$, then $\mathcal{M}$ will be PSD.

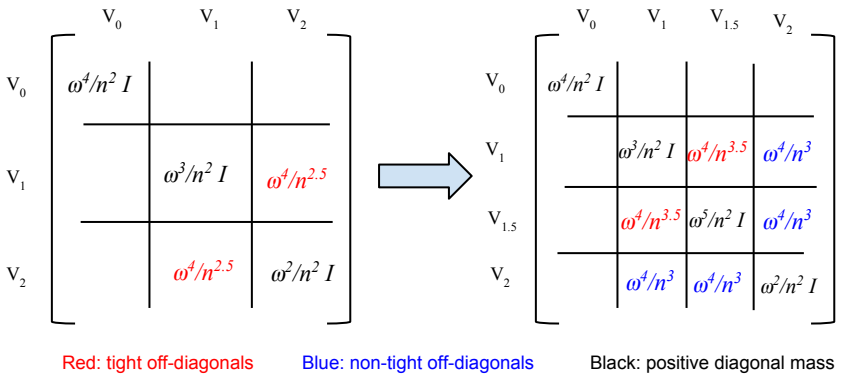


Figure 4.1: The block matrix / subspace decomposition view, before and after the correction. Uninteresting entries left empty.

Because of the dependencies in the random matrix $\mathcal{M}$, the deviation from expectation varies according to the eigenspace. Thus, in [43], the deviation from the expectation is analyzed by first decomposing along the eigenspaces V_0, V_1, V_2.

A second technical idea is required to carry out this decomposition. Because

of the symmetries present in the spaces V_0, V_1, V_2, this decomposition is *very nearly the same* as splitting up the matrix $\mathcal{M}$ in an ostensibly unrelated way. Each entry $\mathcal{M}(I, J)$ of $\mathcal{M}$ for $I, J \in \binom{[n]}{2}$ is the 0/1 indicator for the presence of a clique on $I \cup J$. This indicator is just the AND of all the ± 1 indicators g_b for the presence of the edges $b \in \mathcal{E}(I \cup J)$. Taking a Fourier decomposition of this suggests a way to decompose $\mathcal{M}$ as $\mathcal{M} = \sum_{\text{subsets } S \text{ of edges on 4 nodes}} \mathcal{M}_S$,[6] where the matrix $\mathcal{M}_S$ corresponds to the Fourier character S. The matrices $\mathcal{M}_S$ can be matched up to the subspaces V_0, V_1, V_2 in such a way that those matrices with *larger spectral norm* (corresponding to larger deviations from expectation) have subspaces with *smaller eigenvalues* in their kernels!

Pinpointing the failure of the simple moments. The foregoing is missing one subtlety. Some monomial matrices $\mathcal{M}_S$ do not match nicely to a single subspace. Instead they form cross terms: for example, having V_2 in the left kernel but not in the right kernel (not all these matrices need be symmetric). In fact, it is just such a matrix which keeps the simple moments from remaining PSD beyond $\omega \approx n^{1/3}$.

For the four nodes a_1, a_2, b_1, b_2, consider the monomial $g_{a_1,b_1} g_{a_1,b_2}$ and the corresponding matrix $\mathcal{M}_S(\{a_1, a_2\}, \{b_1, b_2\}) \approx (\omega^4/n^4) g_{a_1,b_1} g_{a_1,b_2}$. The entries $\mathcal{M}_S$ do not depend on a_1, and so there are many repeated rows, which creates a much larger spectral norm for $\mathcal{M}_S$ than if it had independent entries. [43] prove the (tight) bound $\|\mathcal{M}_S\| \approx \omega^4/n^{7/2}$. At the same time, it turns out only to have V_2 in its left kernel, not its right one. Appealing to the above picture, in order to have $\omega^4/n^{7/2} \ll \omega^{5/2}/n^2$, we must have $\omega \ll n^{1/3}$.

[6] This is not quite the whole picture, see Section 4.5.1.1.

Analysis of the correction. We make one further observation about the matrix $\mathcal{M}'_S$ from the previous section: its rows are the tensor squares of the ± 1 neighborhood indicator vectors r_i from above. Our fix to the simple moments, described above as adjusting individual pseudo-expectations, amounts roughly to adding to $\mathcal{M}$ the matrix $(\omega^5/n^5)\sum_i (r_i^{\otimes 2})(r_i^{\otimes 2})^\top$. This carves out of V_2 (our worst subspace from an eigenvalue perspective) a new subspaces space $V_{1.5}$ with eigenvalues lower-bounded by $\lambda_{1.5} \approx \omega^5/n \gg \omega^2/n^2$. Now instead of matching the bad matrix $\mathcal{M}_S$ to V_1 and V_2 as a cross term, we can match it to V_1 and $V_{1.5}$ as a cross term. Then we only need $\omega^4/n^{7/2} \ll \sqrt{\lambda_1 \lambda_{1.5}} \approx \omega^4/n^{7/2}$. With some care in the details, the above picture can be made precise.

However, a crucial point is that the matrix $N = (\omega^5/n^5)\sum_i (r_i^{\otimes 2})(r_i^{\otimes 2})^\top$ doesn't satisfy the clique constraints (in that all entries I, J with $I \cup J$ not a clique should be 0). A chunk of our proof goes into analyzing the discrepancy between the matrix N and its zeroed out version. Our analysis here requires the use of new combinatorial tools (Section 4.4.4) combined with the trace moment method.

We call this the $(i,j)^{th}$ quadratic form in what follows. We write $D = D_0 + D'$ where D_0 is the deviation contributed by the $\tilde{\mathbb{E}}_0$ and D', the deviation contributed by our correction $\mathcal{L}$. We first briefly review the [43] approach to analyzing D_0.

Symmetries of Eigenspaces and Tight Analysis of MPW Operator for Higher Degrees. As we have alluded to already, a key technical step in our proof is to show that certain Fourier-decomposed matrices of the form discussed above have some of the subspaces V_0, V_1, V_2 in their kernels. In the analysis of simple moments for degree 4, [43] use explicit entries for canonical forms of eigenvectors in V_0, V_1 to accomplish this. However this

approach hits analytical roadblocks for the analysis in case of higher degrees. Canonical forms for the eigenvectors are hard to pin down explicitly from the literature in algebraic combinatorics.

To mitigate this difficulty, we take a more principled approach to understand the eigenspaces $V_0, V_1, \ldots, V_d \subseteq \mathbb{R}^{\binom{[n]}{d}}$ in terms of their symmetries. Using techniques from basic representation theory of finite groups, we arrive at an explicit family of symmetries that express any vector in $V_i \subseteq \mathbb{R}^{\binom{[n]}{d}}$ as an explicit linear transformation of some vector in $\mathbb{R}^{\binom{[n]}{i}}$. It also shows that any $v \in V_i \subseteq \mathbb{R}^{\binom{[n]}{d}}$ has the form $\langle v, x^{\otimes d} \rangle$ that's essentially the mutilinearization of $(\sum_j x_j)^{d-i} p(x)$ for some p.

A similar approach was utilized heavily by Grigoriev [59] to prove a sum of squares lower bound for the knapsack problem. While for degree 4 either explicit eigenvectors or our approach will work, although the latter takes some more elbow grease, ours is absolutely vital for our tight analysis of the MPW moments for the higher degrees. We hope that such an approach will be useful for proving better (approaching $\omega \approx \sqrt{n}$) integrality gap for Sum of Squares relaxations of higher degree for Planted Clique and other related problems.

The analysis of the MPW operator at higher degrees also presents other new challenges that do not show up in the special case of degree 4 analyzed in [43]. [43] deal with the optimization version of the degree 4 SOS program which could be potentially weaker than the one we analyze here (and thus our lower bound is technically stronger). Working with the "optimization" version simplifies the analysis in [43] a little bit as the matrix $\mathcal{M}$ has entries that only have local dependence on the graph G. We explicitly work with the feasibility version of the degree 4 SOS program and thus, must deal with the additional

complexity of the entries of $\mathcal{M}$ having a global dependence. As in [85], we deal with this situation by separating $\mathcal{M}$ into matrices L and Δ such that L has only local dependence on the graph G. [85] deal with Δ by a simple entrywise bound, however, employing such a bound yields no improvement over the bound proved in [85] for us. It turns out that we have to do a fine grained analysis of the Δ matrix itself by a decomposition for Δ such that each piece is essentially only locally dependent on the graph. Once we have such a decomposition, our ideas from the analysis of L can be extended to that setting as well.

Finally, our argument for analyzing the spectral norms of each of the pieces of encountered in the decompositions also needs to be much more general than in case of [43] to handle higher degrees. For this, we identify a simple combinatorial structure that controls the norm bounds and allows a general hammer for computing the norms of all the matrices that appear in this analysis. Our proofs here are based on the trace power method and build on the combinatorial techniques in [43].

4.0.6 Preview of Technical Toolkit

In this section we give a preview of the key lemmas that allow us to carry out the analyses described thus far. We have simplified some issues for the sake of exposition; details may be found in Section 4.4.

We are concerned with the matrices in the aforementioned Fourier decomposition. Let $B \subseteq [d] \times [d]$ be a bipartite graph on $2d$ vertices. Let Q_B be an $\binom{[n]}{d} \times \binom{[n]}{d}$ matrix with entries

$$Q_B(I, J) = (-1)^{\text{number of } B\text{-edges which are not } G\text{-edges when the left vertex set of } B \text{ is replaced by } I \text{ and the right one with } J}$$

(We are ignoring what happens if $I \cap J \neq \emptyset$.)

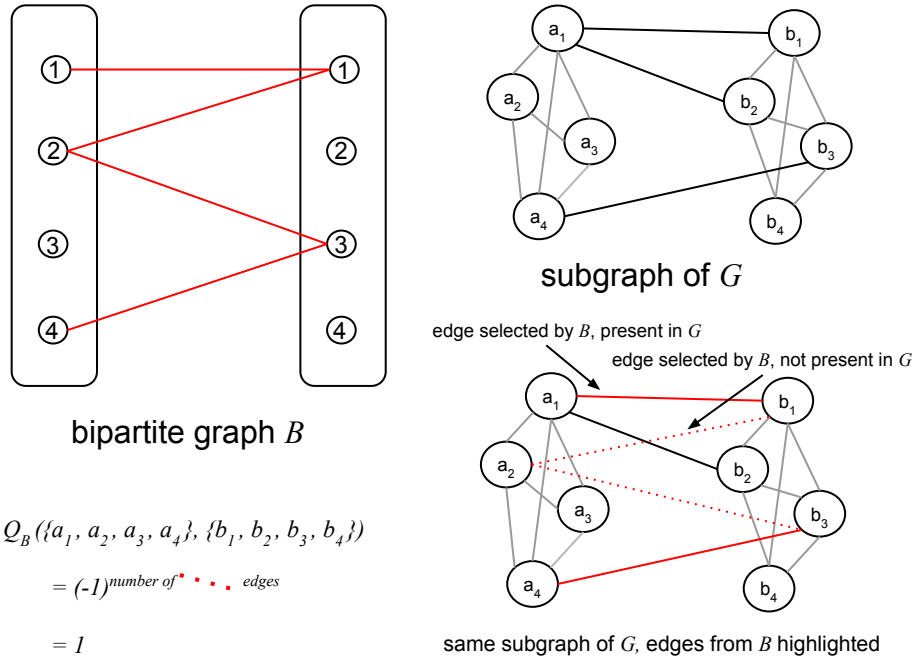


Figure 4.2: Example B and Q_B where f is parity of edges, $d = 4$. Lemma 4.0.4 says that $\Pi_4 Q_B = 0$ and $Q_B r \Pi_4 = Q_B \Pi_3 = 0$. Lemma 4.0.3 says that $\|Q_B\| \approx n^3$ with high probability when $G \sim G(n, 1/2)$, since B contains a 2-matching.

This first lemma bounds the spectral norm of such a matrix in terms of the shape of B.

Lemma 4.0.3 (Informal version of Lemma 4.4.11[7]). *Let c be the number of edges in the maximum matching in B. With high probability, $\|Q_B\| = \tilde{O}(n^{d-c/2})$.*

[7]We use and prove only the cases $c = 1, c = 2$, but the general version follows from almost identical techniques.

We also want to show that these matrices have nontrivial kernels, so we can bound their negative eigenvalues against the parts of the simple moments with larger positive eigenvalues. The following allows us to carry out this matching of Fourier decomposition matrices to eigenspaces $V_0, \ldots, V_d$ of the expectation matrix.

Lemma 4.0.4 (Informal version of Lemma 4.4.9)**.** *Let B_ℓ (B_r) be the subset of vertices on the left (right, respectively) hand side with non zero degrees in B. Let Π_i be the projector to V_i. Then,*

1. *For every $j > |B_\ell|$,*
$$\Pi_j Q_B = 0,$$
2. *For every $i > |B_r|$,*
$$Q_B \Pi_i = 0.$$

The maximum matching cannot be too small when $|B_\ell| + |B_r|$ is large, which allows us to combine these lemmas for every B, either Q_B has small spectral norm or its kernel contains the spaces where the diagonal of the expectation matrix is small.

4.0.7 Related Work

There's a large amount of work on understanding Linear and Semidefinite Programming based hierarchies. A detailed survey on the sum of squares hierarchy and references to works related can be found in [29]. The earliest works on proving SoS lower bounds were due to Grigoriev [59, 60] who showed that degree $\Omega(n)$ SoS does not beat the random assignment for 3SAT or 3XOR even on random instances from a natural distribution. Some of these lower bounds were rediscovered by Schoenebeck [95]. Lower bounds

for SOS essentially rely on gadget reductions from 3SAT or 3XOR and this approach has been understood in some detail [104, 33]. An exception to this methodology is the recent work of Barak et al. in proving SoS lower bounds for pairwise independent CSPs [20]. Even though the lower bounds for CSPs are for random instances, the average-case nature of the problem does not show up as a main analytic issue. There has recently been a surge of interest in understanding the performance of SoS on average-case problems of interest in machine learning, both in proving upper and lower bounds [67, 27, 82, 53, 24].

For the planted clique problem, Feige and Krauthgamer gave an analysis of the performance of the LS+ semidefinite heirarchy tight to within constants [45, 46] giving the state of the art algorithm for finding planted cliques in any fixed polynomial time. Other algorithmic techniques not based on convex relaxations have been studied and shown to fail for planted clique beyond $\omega \approx \sqrt{n}$, most prominantly Markov Chain Monte Carlo (MCMC) [73]. Recently, Feldman et. al. [49] showed a lower bound for (a variant of) the planted clique problem in the restricted class of *statistical algorithms* that generalize MCMC based methods and many other algorithmic techniques. Frieze and Kannan [52] proposed an approach for the planted clique problem through optimizing a degree-3 polynomial related to the random graph. Such polynomials are NP hard to optimize in the worst case but the belief is that the random nature of the polynomials might be helpful. This approach was generalized to higher degree polynomials by Brubaker and Vempala [36].

There has also been recent work on variants of the problem that define Gaussian versions of the planted clique and more generally, the hidden submatrix problems showing, for example, strong inditinguishability results about the spectrum of the associated matrices with and without planting [87]. Finally, the present work builds heavily on in-

dependent papers of Meka, Potechin, and Wigderson [85] and Deshpande and Montanari [43], which we have already thoroughly discussed.

Overview of What Follows. Section 4.1 contains preliminaries. Section 4.2 contains definitions and the necessary background on the simple moments, a.k.a. the MPW operator. Section 4.3 contains the formal definition of our corrected degree 4 moments. Section 4.4 lays out the technical framework for the analyses of the corrected degree 4 moments and for the tightened bounds on the MPW operator at higher degrees. Here we define the Fourier decompositions alluded to above and carry out representation-theoretic arguments about their kernels. Section 4.5 and Section 4.6 use the tools we have built thus far to prove the main theorems. In Section 4.7 we prove a technical concentration result for small subgraphs of $G(n, 1/2)$ required for the analysis. In Section 4.8 we sketch Kelner's argument showing that our analysis of the MPW moments is nearly optimal.

4.1 Preliminaries

We will use the following general notation in the paper.

1. G will denote a draw from $G(n, \frac{1}{2})$ unless otherwise stated.
2. $||x||_2 = ||x||$ denotes the Euclidean 2 norm of a vector $x \in \mathbb{R}^m$.
3. For a square symmetric matrices Q, R, we write $Q \succeq R$ to mean $Q - R$ is positive semidefinite.
4. For any matrix M, $||M||$ denotes its largest singular value, or, equivalently, $||M|| = \max_{x:||x||_2=1} ||Mx||^2$.

5. For matrices M, N of same dimensions, $M \odot N$ denotes their entrywise or *Hadamard* product, i.e., $(M \odot N)(I, J) = M(I, J) \cdot N(I, J)$ for every I, J.

6. For a graph G and any set of two vertices of G, e, g_e denotes the $\{-1, 1\}$ indicator of the edge e being present in G. That is, $g_e = +1$ if e is an edge in G and -1 otherwise.

7. For a set I of vertices of G, $\mathcal{E}(I) = \binom{I}{2}$, the set of all pairs from I.

8. For a pair of subsets of vertices I, J of G, $\mathcal{E}_{ext}(I, J) = \mathcal{E}(I \cup J) \setminus (\mathcal{E}(I) \cup \mathcal{E}(J))$, the set of cut edges between I and J.

9. For a subspace V, Π_V denotes the projector to V.

Fact 4.1.1 (Special Case of Gershgorin Circle Theorem)**.** *For any square matrix $M \in \mathbb{R}^{N \times N}$,*

$$||M|| \leqslant \max_{i \in [N]} \left(\sum_{j=1}^{N} |M_{ij}| \right).$$

The following observations (actually both the same observation in different forms) will come in handy in our analysis.

Lemma 4.1.2. *Let $M \in \mathbb{R}^{n \times n}$ be self-adjoint. Let $W_1, \ldots, W_k$ be an orthogonal decomposition of $\mathbb{R}^n$ into subspaces. Let P_i be the projector to W_i. Let $\lambda_1, \ldots, \lambda_k \geqslant 0$ and suppose for all $i, j \leqslant k$*

$$P_i M P_i \succeq \lambda_i P_i \quad \text{and when } i \neq j \quad \|P_i M P_j\| \leqslant \frac{2}{k} \sqrt{\lambda_i \lambda_j}\,.$$

Then M is PSD.

Proof. Consider a unit vector $x \in \mathbb{R}^n$ and write it as $x = \sum_{i \in [k]} P_i x$. We expand

$$\langle x, Mx \rangle = \sum_{i,j \in [k]} \langle x, P_i M P_j x \rangle \geqslant \sum_i \lambda_i \|P_i x\|^2 - \frac{2}{k} \sum_{i \neq j} \|P_i x\| \|P_j x\| \sqrt{\lambda_i \lambda_i}$$

by our assumptions on $P_i M P_j$ and Cauchy-Schwarz. For each i, j, we know $\frac{1}{k}(\|P_i x\|\lambda_i + \|P_i x\|\lambda_j) \geqslant \frac{2}{k}\|P_i x\|\|P_j x\|\sqrt{\lambda_i \lambda_j}$, which implies that the whole expression is nonnegative. □

Lemma 4.1.3. *Let $M \in \mathbb{R}^{n\times n}$ be self-adjoint. Let V_1, V_2 be subspaces of $\mathbb{R}^n$. Let Π_{V_i} be the projector to V_i. If $\sqrt{\lambda_1\lambda_2} \geqslant \|\Pi_{V_1} M \Pi_{V_2}\|$, then*

$$\Pi_{V_1} M \Pi_{V_2} + \Pi_{V_2} M \Pi_{V_1} \preceq \lambda_1 \Pi_1 + \lambda_2 \Pi_2 \,.$$

Proof. For any $x \in \mathbb{R}^n$ we have

$$\begin{aligned} 2\langle x, \Pi_{V_1} M \Pi_{V_2} x\rangle &\leqslant 2\|\Pi_{V_1} M \Pi_{V_2}\| \cdot \|\Pi_{V_1} x\| \cdot \|\Pi_{V_2} x\| \\ &\leqslant 2(\sqrt{\lambda_1}\|\Pi_{V_1} x\|)(\sqrt{\lambda_2})\|\Pi_{V_1} x\|) \\ &\leqslant \lambda_1\|\Pi_{V_1} x\|^2 + \lambda_2\|\Pi_{V_2} x\|^2 \\ &= \lambda_1\langle x, \Pi_{V_1} x\rangle + \lambda_2\langle x, \Pi_{V_2} x\rangle \,. \end{aligned}$$ □

4.2 The MPW Operator

In this section, we describe the linear operator $\tilde{\mathbb{E}} : \mathbb{P}^n_{2d} \to \mathbb{R}$ for every $d \leqslant O(\log n)$ used by [85] and [43]. It is this operator which we will show gives an integrality gap for degree-$2d$ SoS when $\omega \ll n^{1/(d+1)}$, and it will also form the basis for our improved integrality gap witness at degree four.

The main task in such a setting is to show that $\tilde{\mathbb{E}}$ is positive semidefinite. $\tilde{\mathbb{E}}$ is same as the operator used by [85] who showed that for $\omega \leqslant \Theta(n^{\frac{1}{2d}})$ and graph G drawn at random from $G(n, \frac{1}{2})$, $\tilde{\mathbb{E}}$ is a degree d pseudo-expectation that satisfies all the constraints with high probability over the draw of G. In other words, they showed that $\tilde{\mathbb{E}}$ is a

'cheating" solution that "thinks" that a random graph has a clique of size $\sim n^{1/2d}$ with high probability.

For any set $I \subseteq [n]$, let x_I be the monomial $\Pi_{i \in I} x_i$.

Definition 4.2.1 (MPW operator for clique size ω)**.** For a graph G on n vertices and a parameter $\omega > 0$ we define a linear functional $\tilde{\mathbb{E}} : \mathbb{P}^n_{2d} \to \mathbb{R}$. To describe $\tilde{\mathbb{E}}$ it is enough to describe its values on every monomial x_I for $I \subseteq [n]$, $|I| \leqslant d$. Towards this goal, for any set $I \subseteq [n]$, $|I| \leqslant d$, we define

$$\deg_G(I) = \left|\{S \subseteq [n] : I \subseteq S, |S| = 2d, S \text{ is a clique in } G\ \}\right|.$$

Further, we set $C_{2d} = C_{2d}(G)$ be the number of $2d$-cliques in G.

For every $I \subseteq [n]$, we define:

$$\tilde{\mathbb{E}}[x_I] = \frac{\deg_G(I)}{C_{2d}} \cdot \frac{\binom{\omega}{|I|}}{\binom{2d}{|I|}}. \tag{4.2.1}$$

Our definition of $\tilde{\mathbb{E}}$ is the same as the one used in [85] up to normalization (we explicitly satisfy the normalization condition $\tilde{\mathbb{E}}[1] = 1$). When ω is chosen so that $\tilde{\mathbb{E}}$ is PSD, we often call it the MPW pseudo distribution.

It is easy to check that the linear operator $\tilde{\mathbb{E}}$ satisfies the constraints in (4.0.1) which we record as the following fact:

Fact 4.2.2. *For any graph G, $\tilde{\mathbb{E}}$ defined by Definition 4.2.1 satisfies the constraints described in (4.0.1).*

The main task then is to show that $\tilde{\mathbb{E}}$ is PSD for appropriate range of ω. This task is simplified by another observation from [85] that we state next.

Fact 4.2.3 (Corollary 2.4 in [85]). *For $\tilde{\mathbb{E}}$ of degree d defined in Definition 4.2.1, $\tilde{\mathbb{E}}$ is positive semidefinite iff $\tilde{\mathbb{E}}[p^2] \geqslant 0$ for every multilinear, homogeneous polynomial p of degree d.*

We define a matrix $\mathcal{M} \in \mathbb{R}^{\binom{[n]}{d} \times \binom{[n]}{d}}$ such that $I, J \in \binom{[n]}{d}$,

$$\mathcal{M}(I, J) = \deg_G(I \cup J) \frac{\binom{\omega}{|I \cup J|}}{\binom{2d}{|I \cup J|}}. \tag{4.2.2}$$

Then, from the fact above, showing that $\tilde{\mathbb{E}}$ is PSD is equivalent to proving that $\frac{1}{C_d}\mathcal{M}$ is PSD. The goal of the next section is to establish that with high probability over the draw of $G \sim G(n, 1/2)$, $\mathcal{M}$ is PSD for $\omega \leqslant \tilde{O}(n^{\frac{1}{d+1}})$. This immediately also shows that $\frac{1}{C_d}\mathcal{M}$ is PSD with high probability completing the proof.

Theorem 4.2.4. *With probability at least $1 - 1/n$ over the draw of $G \sim G(n, \frac{1}{2})$ and $d = o(\sqrt{\log(n)})$ $\mathcal{M}$ defined by Definition 4.2.2 is PSD whenever $\omega \leqslant \tilde{O}(n^{\frac{1}{d+1}})$.*

Our analysis improves upon the analysis in [85] and generalizes the improved analysis for the special case of $d = 2$ done in [43]. By a generalization of the counter example due to Kelner, our analysis can actually be shown to be tight. We defer the details of the counter example to the full version. In the remaining part of this section, we begin the task of proving Theorem 4.2.4 by introducing certain simplifications and computing the eigen values of the expected value of the matrix under $G(n, \frac{1}{2})$.

4.2.1 Reduction to PSDness of $\mathcal{M}'$

The presence of some zero rows in $\mathcal{M}$ (corresponding to index sets S that are not cliques in G) poses a problem in analyzing its spectrum. As in [85], we evade this issue by

working with $\mathcal{M}' \in \mathbb{R}^{\binom{[n]}{d} \times \binom{[n]}{d}}$ obtained by filling in the zero rows of $\mathcal{M}$ while not affecting the non zero rows of $\mathcal{M}$. Since the non zero part of $\mathcal{M}$ (for any G) is a sub matrix of $\mathcal{M}'$, proving PSDness of $\mathcal{M}'$ is enough. We describe $\mathcal{M}'$ next and begin by setting up some notation towards that goal:

For any $0 \leqslant i \leqslant d$, let

$$\beta(i) = \frac{\binom{\omega}{2d-i}}{\binom{2d}{2d-i}}.$$

Definition 4.2.5 ("Filled-in" matrix)**.** Let $\mathcal{M}_T \in \mathbb{R}^{\binom{[n]}{d} \times \binom{[n]}{d}}$ be defined by $\mathcal{M}_T(I, J) = \beta(|I \cap J|)$ whenever $I \cup J \subseteq T$ and $\mathcal{E}(T) \setminus (\mathcal{E}(I) \cup \mathcal{E}(J)) \subseteq E$. We define the filled in matrix $\mathcal{M}'$ as:

$$\mathcal{M}' = \sum_{T:|T|=2d} \mathcal{M}_T.$$

Observe that for any I, J, $\mathcal{M}'(I, J)$ is chosen so as to depend only on the edges with one end point in I and the other in J. Intuitively, this corresponds to thinking of I and J as being cliques in G by addition of some edges. Moreoever, $\mathcal{M}'(I, J)$ is chosen so that $\mathcal{M}'(I, J) = M(I, J)$ whenever I, J are actually cliques in G. Thus, as noted above, we have the following fact (which is Lemma 5.1 in [85]).

Fact 4.2.6 (Lemma 5.1 in [85])**.** *$\mathcal{M}$ is PSD if $\mathcal{M}'$ is PSD.*

To analyze $\mathcal{M}'$ we decompose into two parts initially writing $\mathcal{M}' = E + D$ where $E = \mathbb{E}_{G \sim G(n, \frac{1}{2})}[\mathcal{M}']$. We show that E is PSD with all eigenvalues bounded away from 0 in the next subsection following which we analyze the deviation $D = \mathcal{M}' - E$ by writing it as a sum of various pieces and decomposing the action of each piece along the eigenspaces of E in Section 4.5.

Thus, the following lemma completes the proof of Theorem 4.2.4.

Lemma 4.2.7 ($\mathcal{M}'$ is PSD)**.** *With probability at least* $1 - 1/n$ *over the draw of* $G \sim G(n, \frac{1}{2})$ *for* $d = o(\sqrt{\log(n)})$, $\mathcal{M}'$ *defined by Definition 4.2.5 satisfies* $\mathcal{M}' \geq 0$ *whenever* $\omega \leqslant \tilde{O}(n^{\frac{1}{d+1}})$.

4.2.2 The Expectation Matrix

The minimum eigen values of the expectation matrix $E = \mathbb{E}[\mathcal{M}']$ was analyzed in [85] via known results about the Johnson scheme matrices. The same proof also yields all the eigenvalues of E which we note here.

We first describe the entries of the matrix $\mathcal{M}'$.

Fact 4.2.8 (Entries of $\mathcal{M}'$, see Claim 7.3 in [85])**.** *For every* $I, J \in \binom{[n]}{d}$ *and* $E = \mathbb{E}[M']$,

$$E(I,J) = \binom{n - |I \cup J|}{2d - |I \cup J|} \cdot \frac{\binom{\omega}{|I \cup J|}}{\binom{2d}{|I \cup J|}} \cdot 2^{-d^2 - \binom{|I \cap J|}{2}}.$$

Next, we need a basic fact about the (shared) eigenspaces of all the set symmetric matrices, in particular, their number and dimensions which follows from the following well known result from classical theory of Johnson schemes.

Fact 4.2.9 (Lemma 6.6 of [85])**.** *Fix* $n, d \leqslant n/2$ *and let* $\mathcal{J} = \mathcal{J}(n,d)$ *be the set of all set symmetric matrices in* $\mathbb{R}^{\binom{[n]}{d} \times \binom{[n]}{d}}$. *Then, there exist subspaces* $V_0, V_1, \ldots, V_d \in \mathbb{R}^{\binom{[n]}{d}}$ *that are orthogonal to each other such that:*

1. $V_0, V_1, \ldots, V_d$ *are eigen spaces for every* $J \in \mathcal{J}$ *and are isomorphic to distinct irreducible representations of the symmetric group* $\S_n$ *(See Section 4.4 for definitions).*
2. *For* $0 \leqslant j \leqslant d$, $dim(V_j) = \binom{n}{j} - \binom{n}{j-1}$.

Using a nice basis for the matrices in $\mathcal{J}$, one can obtain the following estimates of the eigenvalues of E on V_i for each $0 \leqslant i \leqslant d$:

Lemma 4.2.10 (Eigenvalues of E)**.** *Let $\omega < \frac{n-2d}{3d2^{d-1}}$ and $d \leqslant \omega/2$. Let $\lambda_j(E)$ be the eigenvalue of E on V_j as defined in Fact 4.2.9. Then,*

$$\lambda_j(E) \geqslant \frac{1}{2} \cdot \binom{n-2d+j}{j} \cdot \frac{\binom{\omega}{2d-j}}{\binom{2d}{2d-j}} \cdot 2^{-d^2-\binom{j}{2}} \cdot \binom{n-t-j}{d-t} \cdot \binom{d-j}{t-j} \geqslant 2^{-O(d^2)} \cdot n^d \cdot \omega^{2d-j}.$$

4.3 The Corrected Operator for Degree Four

In this section, we present the pseudodistribution that we will use to show an almost optimal lower bound on the degree 4 SOS algorithm. Our pseudodistribution is obtained by "correcting" the one described in the previous section. The correction itself is inspired by an explicit polynomial described by Kelner who showed that the pseudodistribution from the previous section for degree 4 does not satisfy positivity for $\omega \gg n^{1/3}$.

We now lay some groundwork for defining our modified operator. In the following, we will always work on a fixed graph G on $[n]$ and use $\tilde{\mathbb{E}}_0$ to denote the MPW pseudoexpectation operator for $d = 2$. We start by defining a specific neighborhood indicator vector for every vertex in G. For a vertex $s \in G$, let the vector $r_s \in \mathbb{R}^n$ be given by

$$r_s(j) = \begin{cases} 1 & \text{if } s \sim j \\ -1 & \text{if } s \nsim j \\ 0 & \text{if } s = j \end{cases}.$$

Next, we define the additive correction $\mathcal{L}$ to the MPW pseudoexpectation operator $\tilde{\mathbb{E}}$. $\mathcal{L}$ will be a linear operator on the space of homogeneous degree 4 polynomials. Because of linearity, it is enough to define Ł on the basis of all monomials of degree 4.

Definition 4.3.1 (Correction Term)**.** Let $\gamma > 0$ be a real parameter to be chosen later. Let $\mathcal{L}$ be the linear operator on the linear space of homogeneous, multilinear polynomials of degree 4 such that:

$$\mathcal{L}[x_i x_j x_k x_\ell] = \begin{cases} \gamma(\frac{\omega}{n})^5 \sum_s r_s(i) r_s(j) r_s(k) r_s(\ell) & \text{if } i, j, k, \ell \text{ form a clique in } G \\ 0 & \text{otherwise} \end{cases}.$$

The following is easy to prove.

Fact 4.3.2. *For $\omega > 0$ and $c \ll \omega$, there exists x, $x = \omega \pm O(c/\omega^3)$, such that $\binom{x}{4} = \binom{\omega}{4} + c$.*

We now go on to define the corrected moments $\tilde{\mathbb{E}} : \mathbb{P}_4^n \to \mathbb{R}$.

Definition 4.3.3 (Corrected Pseudoexpectation)**.** We first use the correction operator $\mathcal{L} = \mathcal{L}_\gamma$ to define the corrected moments on all multilinear monomials of degree 4. For every $S \subseteq [n]$, $|S| = 4$, we set:

$$\tilde{\mathbb{E}}[x_S] = \underset{0}{\tilde{\mathbb{E}}}[x_S] + \mathcal{L}[x_S]\,.$$

Next, we want to extend $\tilde{\mathbb{E}}$ to all the monomials so that $\tilde{\mathbb{E}}[1] = 1$ and $\tilde{\mathbb{E}}[\sum_i x_i] = \omega'$ for some $\omega' \approx \omega$. Towards this, we let $c = \sum_{S : S \text{ is a 4-clique in } G} \mathcal{L}[x_S]$. Then, observe that:

$$\tilde{\mathbb{E}} \sum_{S : S \text{ is a 4-clique in } G} x_S = \sum_{S : S \text{ is a 4-clique in } G} \underset{0}{\tilde{\mathbb{E}}}\, x_S + \mathcal{L} x_S = \binom{\omega}{4} + c$$

Then we know there exists $\omega' = \omega \pm O(c/\omega^3) > 0$ satisfying $\binom{\omega'}{4} \overset{\text{def}}{=} \binom{\omega}{4} + c$ (using Fact 4.3.2). Thus, the degree 4 moments we defined "think" $\sum_i x_i = \omega'$. We use this relationship to extend the definition to all the monomials. For every $S \subseteq [n]$, $|S| = 3$,

$$\tilde{\mathbb{E}}[x_S] = \frac{1}{\omega' - 3} \sum_{\ell \notin S} \tilde{\mathbb{E}}[x_{S \cup \ell}]\,.$$

Similarly, for each S: $|S| = 2$,

$$\tilde{\mathbb{E}}[x_S] = \frac{1}{\omega' - 2} \sum_{\ell \notin S} \tilde{\mathbb{E}}[x_{S \cup \ell}] \,.$$

Finally, we set

$$\tilde{\mathbb{E}}[x_i] = \frac{1}{\omega' - 1} \sum_{\ell \neq i} \tilde{\mathbb{E}}[x_i x_\ell]$$

and $\tilde{\mathbb{E}}\, 1 = 1$.

Theorem 4.3.4 (Theorem 1.2.1, formal)**.** *Let $G \sim G(n, 1/2)$. There is $\omega = \Omega(\sqrt{n}/\operatorname{polylog} n)$ so that with probability $1 - 1/n$ the operator $\tilde{\mathbb{E}}$ of Definition 4.3.3 is a valid a degree-4 pseudo-expectation satisfying (4.0.1) for $d = 2$.*

It is not hard to show that $\tilde{\mathbb{E}}$ of Definition 4.3.3 satisfies the constraints in (4.0.1) and the correction above doesn't change ω by a lot. We defer the proofs to Section 4.3.1.

Lemma 4.3.5. *Let $\tilde{\mathbb{E}}$ be the degree-4 corrected moments for clique size ω (Definition 4.3.3). Then, there is ω' such that $\tilde{\mathbb{E}}$ satisfies*

$$\{x_i^2 = x_i\}_{i \in [n]} \,, \quad \{x_i x_j = 0\}_{i \not\sim j \text{ in } G} \,, \quad \{\sum_i x_i = \omega'\} \,.$$

Furthermore, if $G \sim G(n, 1/2)$, then with probability $1 - O(n^{-25})$, $|\omega' - \omega| < O(\gamma \log(n)^2 \omega^2 / n^{5/2})$.

Thus, to show Theorem 4.3.4, the remaining task is to show that $\tilde{\mathbb{E}}$ satisfies positive semidefiniteness. Using Fact Fact 4.2.3, it is enough to show that $\tilde{\mathbb{E}}[p^2] \geqslant 0$ for every homogeneous, multilinear polynomial p of degree 2. This is equivalent to showing that the matrix $\mathcal{N}' \in \mathbb{R}^{\binom{[n]}{2} \times \binom{[n]}{2}}$ defined by $\mathcal{N}'(I, J) = \tilde{\mathbb{E}}[x_{I \cup J}]$ is PSD. Thus, to complete the

proof of Theorem Theorem 4.3.4 we will show the following lemma which is the most technical part of the proof.

Lemma 4.3.6. *There is $\omega_0 = \Omega(\sqrt{n}/\operatorname{polylog} n)$ and $\gamma = \Theta(1)$ so that for $\omega \leqslant \omega_0$, with probability at least $1 - 1/n$ over the draw of $G \sim G(n, \frac{1}{2})$, $\mathcal{N}' \succ 0$.*

4.3.1 Technical Lemmas and Proofs

We proceed here to show that $\tilde{\mathbb{E}}$ from Definition 4.3.3 satisfies the appropriate constraits. We will need the following lemma giving concentration for certain scalar random variables, including the extent to which the correction changes the (pseudo)-expected clique size when $G \sim G(n, 1/2)$.

Lemma 4.3.7. *Let $G \sim G(n, 1/2)$. Let the vectors $r_s \in \mathbb{R}^n$ be as in Definition 4.3.3. There is a universal constant C so that with probability $1 - O(n^{-25})$,*

1. $$\left| \sum_s \sum_{i,j,k,\ell \text{ a clique}} r_s(i) r_s(j) r_s(k) r_s(\ell) \right| \leqslant C n^{5/2} \log(n)^2 .$$

2. *For every i, j, k distinct,*
$$\left| \sum_s \sum_{\ell \text{ in a clique with } i,j,k} r_s(i) r_s(j) r_s(k) r_s(\ell) \right| \leqslant C n \log(n) .$$

3. *For every i, j distinct and every s,*
$$\left| \sum_{k \text{ in a clique with } i,j} r_s(i) r_s(k) \right| \leqslant C \sqrt{n \log n} .$$

Proof. We prove the first item; the others are similar.

The proof is by several applications of McDiarmid's inequality. By a standard Chernoff bound there is a universal constant C_0 so that for every $s \in [n]$ and every $i, j, k \in [n]$, with probabilty $1 - O(n^{-40})$,

$$\left| \sum_{\ell \text{ in a clique with } i,j,k} r_s(\ell) \right| \leqslant C_0 \sqrt{n \log n} \, .$$

Call E_1 the event that this occurs for every s, i, j, k. Clearly $\mathbb{P}(E_1) \geqslant 1 - O(n^{-36})$.

Now for every $s, i, j \in [n]$, we apply McDiarmid's inequality to $|\sum_{k,\ell \text{ in a clique with } i,j} r_s(k) r_s(\ell)|$. We truncate to get rid of the bad event $\neg E_1$. For a graph G, let $f(G) = \sum_{k,\ell \text{ in a clique with } i,j} r_s(k) r_s(\ell)$ if E_1 occurs for G and $f(G) = 0$ otherwise. Now consider any pair of graphs G, G' differing on a single edge (u, v). It is straightforward to show that if $\{u, v\} \cap \{s, i, j\} = \emptyset$ then $|f(G) - f(G')| = O(1)$, while otherwise

$$|f(G) - f(G')| \leqslant \left| \sum_{\ell \text{ in a clique with } i,j,k} r_s(\ell) \right| \leqslant C_1 \sqrt{n \log n}$$

for some other universal constant C_1. So by McDiarmid's inequality there is C_2 so that with probability $1 - O(n^{-34})$,

$$\left| \sum_{k,\ell \text{ in a clique with } i,j} r_s(k) r_s(\ell) \right| \leqslant C_2 n \log n \, .$$

By a similar argument there is C_3 so that for every $s, i \in [n]$, with probability $1 - O(n^{-30})$,

$$\left| \sum_{j,k,\ell \text{ in a clique with } i} r_s(j) r_s(k) r_s(\ell) \right| \leqslant C_3 n^{3/2} (\log n)^{3/2} \, .$$

Let E_2 be the event that this bound holds for every $s, i \in [n]$. So $\mathbb{P}(E_2) \geqslant 1 - O(n^{-27})$. Then, letting $f'(G) = \sum_s \sum_{i,j,k,\ell \text{ a clique}} r_s(i) r_s(j) r_s(k) r_s(\ell)$ if E_2 occurs for G and 0 otherwise, we get that on graphs G, G' differing on an edge (u, v)

$$|g(G) - g(G')| \leqslant C_4 n^{3/2} \log(n)^{3/2}$$

for some other constant C_4. The result follows by a final application of McDiarmid's inequality (we lose a factor of n at this step as opposed to the $\sqrt{n}$ at previous steps because there are $\approx n^2$ edges to be revealed). □

We can now complete the proof of Lemma 4.3.5 using Lemma 4.3.7 and Fact 4.3.2.

Proof of Lemma 4.3.5. The functional $\tilde{\mathbb{E}}$ satisfies the constraints $\{x_i^2 = x_i\}_{i \in [n]}, \{x_i x_j = 0\}_{i \not\sim j \text{ in } G}$ by construction. Let ω' be as in Definition 4.3.3. It is routine to check that for $p(x)$ homogeneous of degree 1, 2, or 3 that $\tilde{\mathbb{E}}\, p(x) \sum_i x_i = \omega' \tilde{\mathbb{E}}\, p(x)$ by definition, so it will be enough to check that $\tilde{\mathbb{E}} \sum_i x_i = \omega'$.

Recall that ω' satisfies $\omega'(\omega' - 1)(\omega' - 2)(\omega' - 3) = 4! \cdot \tilde{\mathbb{E}} \sum_{i,j,k,\ell \text{ a 4 clique}} x_i x_j x_k x_\ell$. Now we expand:

$$\begin{aligned}
\tilde{\mathbb{E}} \sum_i x_i &= \tilde{\mathbb{E}} \frac{1}{\omega' - 1} \sum_i \sum_{j \neq i} x_i x_j \\
&= \tilde{\mathbb{E}} \frac{1}{(\omega' - 1)(\omega' - 2)} \sum_{\substack{i,j,k \\ \text{all distinct}}} x_i x_j x_k \\
&= \frac{1}{(\omega' - 1)(\omega' - 2)(\omega' - 3)} \sum_{\substack{i,j,k,\ell \\ \text{all distinct}}} x_i x_j x_k x_\ell \\
&= \frac{1}{(\omega' - 1)(\omega' - 2)(\omega' - 3)} \cdot 4! \cdot \tilde{\mathbb{E}} \sum_{i,j,k,\ell \text{ a 4 clique}} x_i x_j x_k x_\ell
\end{aligned}$$

$$= \omega' .$$

It remains just to show our claim on $|\omega - \omega'|$. By our choice of ω' and the guarantees of Fact 4.3.2, we get that $|\omega - \omega'| \leqslant |\mathcal{L} \sum_{i,j,k,\ell \text{ a 4-clique}} x_i x_j x_k x_\ell|/\omega^3$ where $\mathcal{L}$ is the correction operator from Definition 4.3.1. By Lemma 4.3.7, this is with probability $1 - O(n^{-25})$ at most $O(\gamma\omega^2 \log(n)^2/n^{5/2})$ when $G \sim G(n, 1/2)$. □

In Section 4.4, we develop some general tools for analyzing the matrices that we encounter before going on to prove Theorem 4.2.4 and Lemma 4.3.6.

4.4 Tools

In this section, we build some general purpose tools helpful in the analysis of the matrices of interest to us. We give statements and proofs that are more or less independent of the rest of the paper with an eye towards future work on planted clique and related problems where one deals with random matrices with structure dependencies. The first three sections focus on building an understanding of the symmetries of the eigenspaces $V_0, V_1, \ldots, V_d$ of set symmetric matrices on $\mathbb{R}^{\binom{[n]}{d} \times \binom{[n]}{d}}$. The last section uses moment method with some combinatorial techniques to obtain tight estimates of spectral norm for certain random matrices with dependent entries.

4.4.1 Background on Representations of Finite Groups

We provide background in the required tools from basic representation theory below.

Definition 4.4.1 (Representation)**.** For a finite dimensional complex vector space V, let

$Hom(V, V)$ be the set of all linear maps from V into V. For any finite group G and $\pi : G \to Hom(V, V)$, the pair (π, V) is said to be a representation of G if π satisfies, for any $g_1, g_2 \in G$,

$$\pi(g_1 \cdot g_2) = \pi(g_1) \cdot \pi(g_2),$$

where the "$\cdot$" on the LHS corresponds to the group operation and on the RHS, the composition of linear maps on V. When the map π is clear from the context (as some natural action of the group G on V), we abuse notation and just say that V is a representation of G.

Let (π, V) be a representation of a group G. A subspace $W \subseteq V$ is said to be a *subrepresentation* if for every $w \in W$, $\pi(g)w \in W$ for every $g \in G$. That is, W is a stable or invariant subspace for all the linear maps $\pi(g)$, one for each $g \in G$. Observe that in this case, (π, W) is another representation of G. A representation (π, V) of G is said to be *irreducible* if for any subspace W invariant under all the linear maps $\pi(g)$ for $g \in G$, $W = V$ or $W = \{0\}$.

Every representation V of G can be decomposed as a direct sum of subspaces each of which is an irreducible representation of G. Further, for any finite group G, there are at most $|G|$ distinct irreducible representation up to isomorphism. For well studied finite groups such as the symmetric group on n elements S_n, the set of irreducible representations are well known and well studied. The power of representation theory in the present context comes from understanding the eigenspace structure of linear operators that are invariant under some action of the group G (in our case $\S_n$).

There is a natural linear action of the permutation group $\S_n$ on $\mathbb{P}_q$ for any q,

denoted by $\pi : \S_n \to Hom(\mathbb{P}_q, \mathbb{P}_q)$: A permutation $\sigma \in \S_n$ when applied to a vector $v \in \mathbb{P}_q$ produces the vector $v' \in \mathbb{P}_q$ such that $v'_I = v_{\sigma(I)}$ for every $I \in \binom{[n]}{q}$. This can be alternately described as multiplication by the permutation matrix associated with σ. Observe that $(\sigma_1 \cdot \sigma_2) \cdot v = \sigma_1 \cdot (\sigma_2 \cdot v)$ and thus, $(\pi, \mathbb{P}_q)$ is a representation of $\S_n$. It is known (See Section 3.2, [16]) that under this action, $\mathbb{P}_d$ can be decomposed as direct sum of subspaces $V_0, V_1, \ldots, V_q$ such that each V_i is an irreducible representation of $\S_n$ and none of V_i, V_j for $i \neq j$ are isomorphic to each other.

The expected moment matrix $E = \mathbb{E}[\mathcal{M}']$ is set symmetric and therefore commutes with the action of $\S_n$ on $\mathbb{P}_d$ described above. This can be easily used to obtain that E has the eigenspaces $V_0, V_1, \ldots, V_d$ discussed above. We need the following consequence of a basic representation-theoretic result.

Fact 4.4.2 (Consequence of Schur's Lemma [98])**.** *Suppose (π, V) and (π', W) are representations of a group G. Suppose $L : V \to W$ is a linear map such that for any $g \in G$ and $v \in V$,*

$$L(\pi(g) \cdot v) = \pi'(g) \cdot L(v).$$

Then, for any irreducible representation $V_i \subseteq V$ under π, $L(V_i) \subseteq W$ is an irreducible representation in W under π'.

4.4.2 Eigenspaces of the Set Symmetric Matrices

We often encounter random $\binom{n}{d} \times \binom{n}{d}$ matrices M indexed by subsets of $[n]$ of size d. For example, a common feature in our setting (as observed in [85]) is that $E = \mathbb{E}[M]$ depends only on $|I \cap J|$.

Definition 4.4.3 (Set Symmetry)**.** A matrix $A \in \mathbb{R}^{\binom{[n]}{d} \times \binom{[n]}{d}}$ is said to be *set symmetric* if for every $S, T, S', T' \in \binom{[n]}{d}$ such that $|S \cap T| = |S' \cap T'|$, $A(S, T) = A(S', T')$.

The set of all set symmetric matrices is known as the *Johnson* scheme in algebraic combinatorics. All such matrices commute and thus share eigenspaces. While the matrices in the Johnson scheme are well studied, the description of the eigenspaces in the literature is hard to use for the purpose of our proofs. We thus take a more direct approach and use basic representation theory in what follows to identify a simple symmetry condition on the eigenspaces of set symmetric matrices which will be useful to understand the spectral properties of the matrices we study.

Lemma 4.4.4. *Let $V_0, V_1, \ldots, V_d \subseteq \mathbb{R}^{\binom{[n]}{d}}$ be the eigenspaces of set symmetric matrices on $\mathbb{R}^{\binom{[n]}{d} \times \binom{[n]}{d}}$ described in the previous section. For any $u \in \mathbb{R}^{\binom{[n]}{t}}$, let $v \in \mathbb{R}^{\binom{[n]}{d}}$ for $d \geqslant t$ be defined so that for each $I \in \binom{[n]}{d}$,*

$$v_I = \sum_{I' \subseteq I, |I'| = t} u_{I'}.$$

Then, $v \in V_0 \oplus V_1 \oplus \cdots \oplus V_t$.

Proof. Let $\mathbb{P}_q$ for any positive integer q be the space of all vectors indexed by elements of $\binom{[n]}{q}$. Consider the standard action of S_n on $[n]$ that sends $i \to \sigma(i)$ for any $\sigma \in \S_n$. This induces a natural action on $\binom{[n]}{q}$ where any $I \in \binom{[n]}{q}$ is sent to $\sigma(I) = \{j \mid \exists i \in I \text{ such that } \sigma(i) = j\}$. This further induces a natural action on $\mathbb{P}_q$ by taking $v = \{v_I\}_{I \in \binom{[n]}{q}}$ and sending it to v' where $v'_I = v_{\sigma^{-1}(I)}$ for every I. A quick check ensures that the action defined above satisfies $\sigma \circ \tau(v) = \sigma(\tau(v))$ for any $\sigma, \tau \in \S_n$. Thus, $\mathbb{P}_q$ is a representation of $\S_n$ under the action defined above for any q. It is easy to check that left multiplication by

any set symmetric matrix from $\mathbb{R}^{\binom{[n]}{q}\times\binom{[n]}{q}}$ commutes with the action of $\S_n$ defined above. From Fact 4.2.9, the eigenspaces of any set symmetric matrix acting on $\mathbb{P}_q$ are given by $V_0, V_1, \ldots, V_q$ such that $\dim(V_i) = \binom{n}{i} - \binom{n}{i-1}$. By Fact 4.2.9 each V_i is isomorphic to distinct irreducible representations of $\S_n$.

Next, consider the map $C : \mathbb{P}_t \to \mathbb{P}_d$ such that for any $u \in \mathbb{R}^{\binom{[n]}{t}}$, the value $C(u) \in \mathbb{R}^{\binom{[n]}{d}}$ is given by v such that $v_I = \sum_{I' \subseteq I, |I'|=t} u_I$. Then, C is linear and we claim that C commutes with the action of $\S_n$ defined above: $\sigma(C(u)) = C(\sigma(u))$. Note that on the LHS, σ refers to the action of $\S_n$ on $\mathbb{P}_d$ while on the RHS, it refers to the action on $\mathbb{P}_t$. We follow the definition to verify this:

$$(\sigma(C(u)))_I = \sigma(\sum_{I' \subseteq I, |I'|=t} u_{I'}) = \sum_{I' \subseteq I, |I'|=t} u_{\sigma^{-1}(I')} = \sum_{I' \subseteq \sigma(I)} u_{I'} = C(\sigma(u))_I.$$

Suppose $u \in V_i^t$ for $i \leqslant t$ where V_i^t is some eigenspace of a set-symmetric $\binom{n}{t} \times \binom{n}{t}$ matrix. Then, by Lemma 4.4.2 $C(V_i^t)$ is an irreducible representation of $\S_n$ and is thus an invariant subspace for the action of $\S_n$ in $\mathbb{P}_d$. By a dimension argument, $C(V_i^t) = V_i$. Thus, $C(u) \in V_i$.

□

4.4.3 Kernels of Patterned Matrices

In this section we design some general tools to understand the spectral structure of matrices that have restricted variations around the set symmetric structure discussed in the previous section. The main tool we will use to establish these results is Lemma 4.4.4 shown in the previous section. Before moving on to this task, we describe a high level

overview of what we intend to do. The following paragraph can be skipped to dive directly into the technical details without the loss of continuity.

The study of the eigenspaces of set symmetric matrices lets us completely understand the spectral structure of the expectation matrix E. In the next section when we analyze the spectrum of $\mathcal{M}'$, we will encounter matrices that depend on the underlying graph G and thus are not set symmetric. However, if the dependence on the underlying graph G is in some sense limited, we hope that some of the nice algebraic properties that set symmetry grants us should perhaps continue to hold. In our case, we will be able to decompose E into various pieces and for each of these pieces, the entry at (I, J) has dependence on the graph G based only on the status (edge vs no edge) of a small number of pairs $(i, j) \in I \times J$. The goal of this section is to develop tools to understand certain (coarse) spectral properties of such matrices.

Our aim is to study matrices in $Q = Q(G) \in \mathbb{R}^{\binom{[n]}{d} \times \binom{[n]}{d}}$ for a graph G on $[d]$ such that $Q(I, J)$ depends on a) the intersections between I and J b) the values of g_b (the edge indicator of G) for pairs b of vertices (from the non intersecting parts of I and J). We first develop some notation to talk about such matrices.

Next, we define *patterns*:

Definition 4.4.5 (Pattern)**.** For $Z_\ell, Z_r \subseteq [d]$, let $\mathcal{B}_{Z_\ell, Z_r}$ be the set of all non-empty bipartite graphs on left and right vertex sets each given by $[d] \setminus Z_\ell$ on the left and $[d] \setminus Z_r$ on the right. Define $\mathcal{B}^q = \cup_{|Z_\ell|, |Z_r| = q} \mathcal{B}_{Z_\ell, Z_r}$. Then, a tuple (B, Z_ℓ, Z_r) for $B \in \mathcal{B}_{Z_\ell, Z_r}$ is said to be a q *pattern* where $q = |Z_\ell| = |Z_r|$. When $q = 0$, we call B itself a pattern.

For any set I, J, consider the "sorting maps" $\zeta_I : [d] \to I$, $\zeta_J : [d] \to J$ i.e, $\zeta_I(1)$

is the least element of I, $\zeta_I(2)$, the next to the least and so on. We can extend ζ_I, ζ_J to subsets of $[d]$ in the natural way. Let B_ℓ (B_r) be the subset of vertices on the left (right) hand side with non zero degrees in $B \in \mathcal{B}_{Z_\ell,Z_r}$. For any $I, J \in \binom{[n]}{d}$, there is a natural map that takes B and obtains a copy of B on vertex sets I and J, via the sorting maps ζ_I and ζ_J from above: $\zeta_{I,J}(B)$ is the bipartite graph on I, J with the edges obtained by taking every edge $b = \{i, j\} \in B$ and adding the edge $\{\zeta_I(i), \zeta_J(j)\}$ to $\zeta_{I,J}(B)$.

We need to understand the effect of applying a permutation $\sigma \in \S_d$ to (B, Z_ℓ, Z_r) for $B \sim \mathcal{B}_{Z_\ell,Z_r}$. Let $\sigma \in \S_d$ be a permutation on $[d]$. Given (B, Z_ℓ, Z_r), σ has two natural actions. The left action of σ on (B, Z_ℓ, Z_r) produces $\sigma \circ (B, Z_\ell, Z_r) \stackrel{\text{def}}{=} (\sigma \circ B, \sigma(Z_\ell), Z_r)$ where each edge $(i, j) \in B$ is sent into $(\sigma(i), j)$ in $\sigma \circ B$. We similarly define the right action of σ on (B, Z_ℓ, Z_r) that produces $(B, Z_\ell, Z_r) \circ \sigma \stackrel{\text{def}}{=} (B \circ \sigma, Z_\ell, \sigma(Z_r))$. Each of these two actions defines a subgroup that leaves B invariant.

Definition 4.4.6 (Automorphism Groups)**.** Let $B \in \mathcal{B}_{Z,Z'}$ be a labeled bipartite graph. We define the left automorphism group of (B, Z_ℓ, Z_r) as

$$Aut_\ell(B, Z_\ell, Z_r) = \{\sigma \in \S_d \mid \sigma \circ (B, Z_\ell, Z_r) = (B, Z_\ell, Z_r)\},$$

and the right automorphism group of B as

$$Aut_r(B, Z_\ell, Z_r) = \{\sigma \in \S_d \mid (B, Z_\ell, Z_r) \circ \sigma = (B, Z_\ell, Z_r)\}.$$

Next, we define equivalence classes of the patterns (B, Z_ℓ, Z_r).

Definition 4.4.7 (Similar Patterns)**.** For patterns (B, Z_ℓ, Z_r) is *left* similar to (B', Z'_ℓ, Z'_r) and write $(B, Z_\ell, Z_r) \sim_\ell (B', Z'_\ell, Z'_r)$ if there exists a $\sigma \in \S_d$ such that $\sigma \circ (B, Z_\ell, Z_r) =$

(B', Z'_ℓ, Z'_r). Similarly, we say that (B, Z_ℓ, Z_r) is *right* similar to (B', Z'_ℓ, Z'_r) and write $(B, Z_\ell, Z_r) \sim_r (B', Z'_\ell, Z'_r)$ if there exists a $\sigma \in \S_d$ such that $(B', Z'_\ell, Z'_r) = (B, Z_\ell, Z_r) \circ \sigma$.

We are now ready to define patterned matrices:

Definition 4.4.8 (Patterned Matrices)**.** Let (B, Z_ℓ, Z_r) be a q-pattern. Let $f : \{-1, 1\}^B \to \mathbb{R}$ be a function that maps a $\{-1, 1\}$ labeling of the pairs in B to $\mathbb{R}$. For a graph G on $[n]$ vertices, the patterned matrix with pattern (B, Z_ℓ, Z_r) defined by f is a matrix in $Q = Q_{B,Z_\ell,Z_r,f}(G) \in \mathbb{R}^{\binom{[n]}{d} \times \binom{[n]}{d}}$ such that

$$Q(I, J) = \begin{cases} f(\{g_b\}_{b \in \zeta_{I,J}(B)}), \text{ for every } I, J \ \zeta_I(Z_\ell) = \zeta_J(Z_r) \\ 0, \text{ otherwise.} \end{cases}$$

When $q = 0$, we write $Q_{B,f}$ for the corresponding patterned matrix.

The following result describes the kernels of certain symmetrized sums of $Q_{B,Z,f}$ and is the main claim of this section.

Lemma 4.4.9. *For graph G, a q-pattern (B, Z_ℓ, Z_r) and $f : \{-1, 1\}^B \to \mathbb{R}$, let $Q = Q_{B,Z_\ell,Z_r,f}(G) \in \mathbb{R}^{\binom{[n]}{d} \times \binom{[n]}{d}}$ be the corresponding patterened matrix. Define the left and right symmetrized version of Q by:*

$$Q^\ell = \sum_{(B', Z'_\ell, Z'_r) \sim_\ell (B, Z_\ell, Z_r)} Q_{B', Z'_\ell, Z'_r, f},$$

and

$$Q^r = \sum_{(B', Z'_\ell, Z'_r) \sim_r (B, Z_\ell, Z_r)} Q_{B', Z_\ell, Z_r, f},$$

respectively. Let B_ℓ (B_r) be the subset of vertices on the left (right, respectively) hand side with non zero degrees in B. Then,

1. *For every* $j > |B_\ell| + q$,

$$\Pi_j^\dagger Q^\ell_{B,Z,f} = 0,$$

2. *For every* $i > |B_r| + q$,

$$Q^r_{B,Z,f}\Pi_i = 0.$$

Proof. Observe that $Q^\ell_{B,Z_\ell,Z_r,f} = Q^\ell_{B',Z'_\ell,Z'_r,f}$ for any $(B, Z_\ell, Z_r) \sim_\ell (B', Z'_\ell, Z'_r)$. This motivates us to first obtain a more symmetric looking expression for Q^ℓ and Q^r. Let $\S_d/Aut_\ell((B, Z_\ell, Z_r))$ (and correspondingly, $\S_d/Aut_r((B, Z_\ell, Z_r))$) be the group of left (right) cosets of $Aut_\ell((B, Z_\ell, Z_r))$ ($Aut_r((B, Z_\ell, Z_r))$ respectively). We have:

$$Q^\ell = \sum_{(B,Z_\ell,Z_r)\sim_\ell(B',Z'_\ell,Z'_r)} Q_{B',Z'_\ell,Z'_r,f} = \sum_{\tau\in\S_d/Aut_\ell((B,Z_\ell,Z_r))} Q_{\tau\circ(B,Z_\ell,Z_r)} = \frac{1}{|Aut_\ell((B, Z_\ell, Z_r))|}\sum_{\sigma\in\S_d} Q_{\sigma\circ(B,Z_\ell,Z_r)}.$$

Similarly, we have:

$$Q^r_{B,Z_\ell,Z_r,f} = \frac{1}{|Aut_r((B, Z_\ell, Z_r))|}\sum_{\sigma\in\S_d} Q_{(B,Z_\ell,Z_r)\circ\sigma,f}.$$

We now begin the argument for proving the first claim. The second claim has an analogous proof. Consider an arbitrary $v = \{v_I\}_{I\in\binom{[n]}{d}} \in \mathbb{R}^{\binom{[n]}{d}}$. We will show that $Q^\ell v \in V_0 \oplus V_1 \oplus \cdots \oplus V_{q+|B_\ell|}$. Towards this goal, define a vector $u = \{u_T\}_{T\in\binom{[n]}{k}} \in \mathbb{R}^{\binom{[n]}{k}}$ as follows: for each $T \in \binom{[n]}{k}$, let $I_T \in \binom{[n]}{d}$ be arbitrary subject to the constraint that $I_T \supseteq T$. We define:

$$u_T = \sum_{\substack{(B',Z'_\ell,Z'_r)\\ \sim_\ell(B,Z_\ell,Z_r)\\ B'_\ell\cup Z'_\ell = T}} \sum_{J:\zeta_J(Z_r)=\zeta_{I_T}(Z'_\ell)} v_J f(\{g_b\}_{b\in\zeta_{I_T,J}(B')}).$$

We first show that u above is well defined in that the definition does not depend on the specific subset I_T used so long as $I_T \supseteq T$. We adopt the notation (that ignores the "direction" of action) $B^\sigma \stackrel{\text{def}}{=} \sigma \circ B$ only for the calculations that follow.

Claim 4.4.10. Fix $T \in \binom{[n]}{q+|B_\ell|}$ and let $I_1, I_2 \in \binom{[n]}{d}$ such that $T \subseteq I_1, I_2$. Then,

$$\sum_{\substack{(B',Z'_\ell,Z'_r) \\ \sim_\ell (B,Z_\ell,Z_r) \\ B'_\ell \cup Z'_\ell = T}} \sum_{J:\zeta_J(Z_r)=\zeta_{I_1}(Z'_\ell)} v_J f(\{g_b\}_{b\in\zeta_{I_1,J}(B')}) = \sum_{\substack{(B',Z'_\ell,Z'_r) \\ \sim_\ell (B,Z_\ell,Z_r) \\ B'_\ell \cup Z'_\ell = T}} \sum_{J:\zeta_J(Z_r)=\zeta_{I_2}(Z'_\ell)} v_J f(\{g_b\}_{b\in\zeta_{I_2,J}(B')}).$$

Proof of Claim. We can equivalently write the claim above as:

$$\sum_{\sigma\in\S_d} \sum_{J:\zeta_J(Z_r)=\zeta_{I_1}(\sigma(Z_\ell))} v_J \cdot f(\{g_b\}_{b\in\zeta_{I_1,J}(B^\sigma)}) = \sum_{\sigma\in\S_d} \sum_{J:\zeta_J(Z_r)=\zeta_{I_2}(\sigma(Z_\ell))} v_J \cdot f(\{g_b\}_{b\in\zeta_{I_2,J}(B^\sigma)}) \quad (4.4.1)$$

We start with the LHS and observe that for any $\tau \in \S_d$ it equals:

$$\sum_{\sigma\in\S_d} \sum_{J:\zeta_J(Z_r)=\zeta_{I_1}(\sigma\circ\tau(Z_\ell))} v_J \cdot f(\{g_b\}_{b\in\zeta_{I_1,J}(B^{\sigma\circ\tau})}).$$

We show that there exists a τ such that $\zeta_{I_1}(\sigma \circ \tau(Z_\ell)) = \zeta_{I_2}(\sigma(Z_\ell))$ and $\zeta_{I_1,J}(\sigma \circ \tau \circ B) = \zeta_{I_2,J}(\sigma \circ B)$.

For each $i \in I'$, let $b_i^2 \in [d]$ be such that $\zeta_{I_2}(b_i^2) = i$. Similarly, for each $i \in I'$, let $b_i^1 \in [d]$ be such that $\zeta_{I_1}(b_i^1) = i$. For each $i \in I'$, choose τ such that $\tau(b_i^2) = b_i^1$. Then, $\zeta_{I_1}(\tau(b_i^2)) = \zeta_{I_1}(b_i^2) = i$. Thus, $\zeta_{I_1,J}(B'^\tau) = \zeta_{I_2,J}(B')$ for every $(B', Z'_\ell, Z'_r) \sim_\ell (B, Z_\ell, Z_r)$. □

We can now show that $(Q^\ell v)_I = \sum_{I'\subseteq I, |I'|=q+|B_\ell|} u_{I'}$. We now observe:

$$(Q^\ell v)_I = \sum_{J\in\binom{[n]}{d}} Q^\ell(I,J)\cdot v_J \quad (4.4.2)$$

$$= \frac{1}{|Aut_\ell(B,Z_\ell,Z_r)|} \sum_{J\in\binom{[n]}{d}} \sum_{\sigma\in\S_d} Q_{B',Z'_\ell,Z'_r,f} v_J$$

Keeping Js that correspond to non-zero entries in $Q_{B',Z'_\ell,Z_r\,f}$ (4.4.3)

$$= \frac{1}{|Aut_\ell(B, Z_\ell, Z_r)|} \sum_{\sigma \in \S_d} \sum_{J : \zeta_J(Z_r) = \zeta_I(Z'_\ell)} Q_{B', Z'_\ell, Z'_r, f} v_J$$

$$= \frac{1}{|Aut_\ell(B, Z_\ell, Z_r)|} \sum_{\sigma \in \S_d} \sum_{J : \zeta_J(Z_r) = \zeta_I(Z'_\ell)} f(\{g_b\}_{b \in \zeta_{I,J}(B^\sigma)}) v_J$$

Using that $\cup_{I' \subseteq I} \{\sigma \mid \zeta_I(B^\sigma_\ell) \cup \sigma(Z_\ell) = I'\}$ forms a partition of $\S_d$ indexed by I', (4.4.4)

$$= \sum_{I' \subseteq I} \frac{1}{|Aut_\ell(B, Z_\ell, Z_r)|} \sum_{\sigma \in \S_d,\, B^\sigma_\ell \cup \sigma(Z_\ell) = I',\, J : \zeta_J(Z_r) = \zeta_I(Z'_\ell)} f(\{g_b\}_{b \in \zeta_{I,J}(B^\sigma)}) v_J$$

Since each $(B', Z'_\ell, Z'_r) \sim_\ell (B, Z_\ell, Z_r)$ s.t. $B'_\ell \cup Z'_\ell = I'$ (4.4.5)

occur $|Aut_\ell(B, Z_\ell, Z_r)|$ times in inner sum,

$$= \sum_{I' \subseteq I} \sum_{(B', Z'_\ell, Z'_r) \sim_\ell (B, Z_\ell, Z_r),\, B'_\ell \cup Z'_\ell = I'} \sum_{J : \zeta_J(Z_r) = \zeta_I(Z'_\ell)} f(\{g_b\}_{b \in \zeta_I(B'_\ell)}) v_J$$

Using Claim 4.4.10: (4.4.6)

$$= \sum_{I' \subseteq I} u_{I'}. \tag{4.4.7}$$

This completes the proof using Lemma 4.4.4. □

4.4.4 Concentration for Locally Random Matrices over $G(n, \frac{1}{2})$

The goal of this section is to prove strong concentration bounds for the matrices will encounter in our analysis. The first result is a spectral concentration bound for the patterned matrices $Q = Q_{B,f}(G)$ for $G \sim G(n, \frac{1}{2})$ when $f : \{-1, 0, 1\}^B \to \mathbb{R}$ is given by $f(x) = \Pi_{b \in B} x_b$. In other words, the entry $Q(I, J)$ is the product of the edge indicator variables g_b for $b \in \zeta_{I,J}(B)$. These bounds will be used in Section 4.5.

Lemma 4.4.11. *For $d \geqslant 2$, $d = O(\log(n))$, and a bipartite graph $B \in \mathcal{B}$, let $Q = Q_{B,f}$ be a*

patterned matrix with $f(x) = \prod_{b \in B} x_b$. *That is,*

$$Q(I,J) = \begin{cases} \prod_{b \in \zeta_{I,J}(B)} g_b & \text{if } I \cap J = \emptyset \\ 0 & \text{otherwise} \end{cases},$$

Then:

1. *When* B *contains a* 2*-matching, then* $\mathbb{P}(\|Q\| \geqslant n^{d-1}(\log n)^3) \leqslant O(n^{-10})$.
2. *When* B *is not the empty graph,* $\mathbb{P}(\|Q\| \geqslant n^{d-1/2}(\log n)^3) \leqslant O(n^{-10})$.

The next main result of this section considers a different class of matrices that appear in the analysis in Section 4.6.

Lemma 4.4.12. *Let* $U \subseteq [2] \times [2]$ *be a bipartite graph on* 4 *vertices and suppose* U *is nonempty. Let* $M \in \binom{[n]}{2} \times \binom{[n]}{2}$ *be a matrix with the entry at* $\{a_1, a_2\}$, $\{b_1, b_2\}$ *for* $a_1 \leqslant a_2$ *and* $b_1 \leqslant b_2$ *are given by*

$$M[\{a_1,a_2\},\{b_1,b_2\}] = \begin{cases} \sum_{k \in [n]} g_{k,a_1} g_{k,a_2} g_{k,b_1} g_{k,b_2} \prod_{(i,j) \in U} g_{a_i,b_j} & \text{if } |\{a_1,a_2,b_1,b_2\}| = 4 \\ 0 & \text{otherwise} \end{cases},$$

Recall that we set $g_{aa} = 0$ *for every* $a \in [n]$ *by convention. Then, whenever* U *is non-empty,* $\mathbb{P}(\|M\| \geqslant n^{3/2}(\log n)^3) \leqslant O(n^{-10})$. *If* U *is the empty graph, then* $\mathbb{P}(\|M\| \geqslant n^2(\log n)^3) \leqslant O(n^{-10})$.

The proofs of both these results are based on the standard idea of analyzing the trace of higher powers of a matrix to prove bounds on its spectral norm. The proof of Lemma 4.4.11 is similar to the proofs via the trace power method for bounding the norms of matrices as presented in [43]. The general format we present here will come in

handy for multiple applications to various matrices in Section 4.5. Lemma 4.4.12 deals with somewhat more complicated matrices that appear in the analysis of the corrected operator for degree 4 lower bound. Nevertheless, as is common in such proofs, the analysis is based on a combinatorial analysis of the terms that make non zero contribution to the trace powers combined with the simplifying effect of random partitioning based arguments. We describe the details of the proof in the following section.

4.4.4.1 General Tools

Before diving into the details, we present three general purpose tools that we will employ repeatedly in our analyses. For analyzing the spectral norm of a matrix $Q \in \mathbb{R}^{\binom{[n]}{d}\times\binom{[n]}{d}}$, the first tool allows us to analyze instead a related matrix $Q' \in \mathbb{R}^{n^d \times n^d}$. That is, instead of rows and columns being indexed by subsets of vertices as in Q, Q' has rows and columns indexed by ordered tuples of vertices of size d. This transformation is not hard as one can find Q as a principal submatrix of Q'.

Lemma 4.4.13 (Sets to Ordered Tuples)**.** *For any $Q \in \mathbb{R}^{\binom{[n]}{d}\times\binom{[n]}{d}}$ define the matrix $Q' \in \mathbb{R}^{n^d \times n^d}$ such that for any ordered tuple $S = (a_1, a_2, \ldots, a_d), T = (b_1, b_2, \ldots, b_d) \in [n]^d$, $Q'(S,T) = Q(\{a_1, a_2, \ldots, a_d\}, \{b_1, b_2, \ldots, b_d\})$. Then, $\|Q\| \leqslant \|Q'\|$.*

Proof. It is enough to show that Q' occurs as a principal submatrix of Q. For this, take the submatrix of rows and columns of M indexed by tuples $(a_1, \ldots, a_d)$ in sorted order, i.e., with $a_1 \leqslant a_2 \leqslant \ldots a_d$. □

We will use the following lemma to break dependencies in certain random matrices by decomposing them into matrices whose entries, while still dependent, have additional

structure.

Lemma 4.4.14 (Random Partitioning)**.** *For $d \in \mathbb{N}$, let $Q \in \mathbb{R}^{n^d \times n^d}$.*

1. *Suppose $Q(I, J) = 0$ when $I \cap J \neq \emptyset$. Let $(S_1^1, \ldots, S_k^1), \ldots, (S_1^r, \ldots, S_k^r)$ be a sequence of partition of $[n]$ into k bins. Each partition induces a matrix based on Q as follows:*

$$Q_i[(a_1, \ldots, a_d), (b_1, \ldots, b_d)] = \begin{cases} Q[(a_1, \ldots, a_d), (b_1, \ldots, b_d)] & \text{if } a_j, b_j \in S_j^i \text{ for } j < k \\ & \text{and } a_j, b_j \in S_k \text{ for } j \geqslant k \\ & \text{and for all } i' < i,\ M_{i'}[(a_1, \ldots, a_d), (b_1, \ldots, b_d)] = 0 \\ 0 & \text{otherwise} \end{cases}.$$

   Then, there is a family of partitions $(S_1^1, \ldots, S_k^1), \ldots, (S_1^r, \ldots, S_k^r)$ such that $Q = \sum_{i=1}^r Q_i$ with $r \leqslant O(k^k \log n)$.

2. *Let $Q^j \in \mathbb{R}^{n^d \times n^d}$, for each $1 \leqslant j \leqslant n$, be matrices such that $Q^j(I, J) = 0$ whenever $I \cap J \neq \emptyset$ or $j \in I \cup J$. Suppose $Q = \sum_{j=1}^n Q^j$. For a partition $(S_1, \ldots, S_k, T)$ of $[n]$ into $k + 1$ parts, say that $j, (a_1, \ldots, a_d), (b_1, \ldots, b_d)$ respect the partition if $j \in T$, $a_i, b_i \in S_i$ for all i. Let a sequence of partitions $(S_1^1, \ldots, S_k^1, T^1), \ldots, (S_1^r, \ldots, S_k^r, T^r)$ of $[n]$ into $k + 1$ parts induce matrices M_i in the following way.*

$$Q_i[(a_1, \ldots, a_d), (b_1, \ldots, b_d)] = \sum_{j \in T^i_{(a_1, \ldots, a_d), (b_1, \ldots, b_d)}} Q^j[(a_1, \ldots, a_d), (b_1, \ldots, b_d)]$$

   where $T^i_{(a_1, \ldots, a_d), (b_1, \ldots, b_d)}$ is the set of indices j so that $j, (a_1, \ldots, a_d), (b_1, \ldots, b_d)$ respect the partition $(S_1^i, \ldots, S_k^i, T^i)$ and do not respect any partition $(S_1^{i'}, \ldots, S_k^{i'}, T^{i'})$ for any $i' < i$.

Then, there is a family $(S_1^1, \ldots, S_k^1, T^1), \ldots, (S_1^r, \ldots, S_k^r, T^r)$ of partitions of $[n]$ so that $Q = \sum_{i=1}^r Q_i$ with $r \leqslant O((k + 1)^{k+1} \log n)$.

Proof of Lemma 4.4.14. We present the proof of 1; the proof of 2 is almost identical. For r to be chosen later, we pick partitions $(S_1^1, \ldots, S_k^1), \ldots, (S_1^r, \ldots, S_k^r)$ uniformly at random and independently so that each is partition of $[n]$ into sets of size n/k each.

Call $(a_1, \ldots, a_d), (b_1, \ldots, b_d)$ *good* at step i if $a_j, b_j \in S_j^i$ for every $j < k$ and $a_j, b_j \in S_k^i$ if $j \geqslant k$. It is enough to show that after $r \leqslant O(k^k \log n)$ steps the probability that every $\{a_1, \ldots, a_d, b_1, \ldots, b_d\}$ of size $2d$ is good at some step $i \leqslant r$.

Fix some $(a_1, \ldots, a_d), (b_1, \ldots, b_d)$ with $|\{a_1, \ldots, a_d, b_1, \ldots, b_d\}| = 2d$. It is good at step i with probability at least k^{-k}. Since the steps are independent, after r steps

$$\begin{aligned}\mathbb{P}((a_1, \ldots, a_d), (b_1, \ldots, b_d) \text{ is good}) &\geqslant (1 - \tfrac{1}{k^k})^r \\ &= ((1 - \tfrac{1}{k^k})^{k^k})^{r/k^k} \\ &\leqslant (\tfrac{1}{e})^{r/k^k}\end{aligned}$$

which is at most $1/n^{10d}$ for some $r = O(k^k \log n)$.

Taking a union bound over all $O(n^{2d})$ tuples $(a_1, \ldots, a_d), (b_1, \ldots, b_d)$ with $|\{a_1, \ldots, a_d, b_1, \ldots, b_d\}| = 2d$ completes the proof. □

Finally, the following lemma relates the norms of certain matrices in $X \in \mathbb{R}^{\binom{[n]}{d} \times \binom{[n]}{d}}$ that have non zero entry (I, J) only if $|I \cap J| = q$ to a certain *lift* of X that lives in $\mathbb{R}^{\binom{[n]}{d-q} \times \binom{[n]}{d-q}}$ and has non zero entries I, J only when $I \cap J = \emptyset$. The latter case is easier to handle and the idea of lifts helps reducing the norm computation for lifts of X to that of X.

Definition 4.4.15 (Lifts of Matrices, Equation 8.5 in [85])**.** For a matrix $X \in \mathbb{R}^{\binom{[n]}{d-i} \times \binom{[n]}{d-i}}$ for some $0 \leqslant i \leqslant d$ such that $X(I', J') = 0$ whenever $I' \cap J' \neq \emptyset$, define the lift $X^{(i)} \in \mathbb{R}^{\binom{[n]}{d} \times \binom{[n]}{d}}$

to be the matrix defined by:

$$X^{(i)}(I,J) = \begin{cases} X(I \setminus (I \cap J), J \setminus (I \cap J)), \text{ if} |I \cap J| = i \\ 0, \text{ otherwise.} \end{cases}$$

The usefulness of the above definition is captured by the following claim:

Fact 4.4.16 (Lemma 8.4 in [85]). *Let $X \in \mathbb{R}^{\binom{[n]}{d-i} \times \binom{[n]}{d-i}}$ for some $0 \leqslant i \leqslant d$ such that $X(I', J') = 0$ whenever $I' \cap J' \neq \emptyset$. Then, for the lift $X^{(i)}$ of X, we have:*

$$\|X^{(i)}\| \leqslant \binom{d}{i}^2 \cdot \|X\|.$$

4.4.4.2 Graph-Theoretic Definitions and Lemmas

In this section, we set up some notation and definitions helpful in our proofs of the main results of this section. The next few definitions and notation are generalization of the ones used in [43] to general degrees d and are useful in the proof of Lemma 4.4.11.

Definition 4.4.17. Let U be a bipartite graph on vertices $\{1, 2, \ldots, d\} \times \{1', 2', \ldots, d'\}$. A *$U$-ribbon* of length 2ℓ is a graph R on $2\ell d$ vertices

$$a_1^1, \ldots, a_d^1, \ldots, a_1^\ell, a_d^\ell$$
$$b_1^1, \ldots, b_d^1, \ldots, b_1^\ell, b_d^\ell .$$

We install edges in R by placing a copy of U on vertices $1, 2, \ldots, d$ and $1', 2', \ldots, d'$ (with the label i or i' matching the upper index of as and bs respectively) on $a_1^i, \ldots, a_d^i, b_1^{i-1}, \ldots, b_d^{i-1}$ for every $i \leqslant d$. For $i = 0$, we treat $i - 1$ as d (modular addition). Often we will omit the length parameter 2ℓ when it is clear from context.

Definition 4.4.18. Let G be a graph. A *labeled* U-ribbon R is a tuple (R, F) where R is a U-ribbon and $F : R \to G$ is a map labeling each vertex of R with a vertex in G. We require that for (u, v) an edge in R, $F(u) \neq F(v)$.

Definition 4.4.19. Let (R, F) be a labeled U-ribbon where U has $2d$ vertices. We say (R, F) is *disjoint* if for every i,

$$|\{F(a_i^1), \ldots, F(a_i^d), F(b_i^1), \ldots, F(b_i^d)\}| = |\{F(a_{i-1}^1), \ldots, F(a_{i-1}^d), F(b_{i-1}^1), \ldots, F(b_{i-1}^d)\}| = 2d\,.$$

Definition 4.4.20. Let (R, F) be a labeled U-ribbon where U has $2d$ vertices. We say that (R, F) is *contributing* if no element of the multiset $\{(F(u), F(v)) : (u, v) \in R\}$ occurs with odd multiplicity.

The following combinatorial lemma will serve as a tool in the proofs of the main results for this section.

Lemma 4.4.21. *Let (R, F) be a contributing labeled U-ribbon of length 2ℓ. Recall that R has vertex set a_j^i, b_j^i for $i \in \ell$ and $j \in [d]$. Let $k \leqslant d$. Suppose that the sets*

$$\{F(a_1^i), F(b_1^i)\}_{i\in[\ell]}, \ldots, \{F(a_k^i), F(b_k^i)\}_{i\in[\ell]}, \quad \{F(a_j^i), F(b_j^i)\}_{i\in[\ell], j\in[k+1,d]}$$

are disjoint. Then if U contains the edges $\{(1,1), \ldots, (k,k)\}$ (where we identify the vertex set of U with $[d] \times [d]$), $\{F(u) : u \in R\}$ has size at most $(2d - k)\ell + k$.

Proof. The assumption on U implies that R contains the cycles

$$C_1 \stackrel{\text{def}}{=} (a_1^1, b_1^1, \ldots, b_\ell^1, a_1^1)$$

$$\ldots$$

$$C_k \stackrel{\text{def}}{=} (a_1^k, b_1^k, \ldots, b_\ell^k, a_1^k)\,.$$

In order for (R, F) to be contributing, every edge $(u, v) \in R$ must have a partner $(u', v') \neq (u, v)$ so that $F(u') = F(u)$ and $F(v') = F(v)$. By our disjointness assumption, every edge in cycle C_i must be partnered with another edge in C_i. Thus, now temporarily identifying edges when they are labeled identically, each C_i is a connected graph with at most ℓ unique edges (since each of the 2ℓ edges must be partnered). It thus has at most $\ell + 1$ unique vertex labels. Among the cycles $C_1, \ldots, C_k$, there are thus at most $k(\ell + 1)$ unique vertex labels. In the rest of the ribbon R there can be at most $2\ell(d - k)$ unique vertex labels, because once the cycles $C_1, \ldots, C_k$ are removed there are only that many vertices left in R. So in total there are at most $k(\ell + 1) + 2\ell(d - k) = (2d - k)\ell + k$ unique labels. □

The next few definitions and notation are needed in the proof of Lemma 4.4.12.

Definition 4.4.22. Let U be a bipartite graph on vertices a_1, a_2, b_1, b_2. A *fancy U-ribbon* R of length 2ℓ is a graph on vertices $c_1, \ldots, c_{2\ell}, a_1^1, a_1^2, \ldots, a_\ell^1, a_\ell^2, b_1^1, b_1^2, \ldots, b_\ell^1, b_\ell^2$. On the a and b vertices, R restricts to a U-ribbon of length 2ℓ. Additionally, it has edges $(c_i, a_i^1), (c_i, a_i^2), (c_i, b_i^1), (c_i, b_i^2)$.

Where G is a graph, a labeled fancy U-ribbon is a tuple (R, F) where R is a fancy U-ribbon and $F : R \to G$ labels each vertex of R with a vertex in G. We require for any edge $(u, v) \in R$ that $F(u) \neq F(v)$.

Lemma 4.4.23. *Let U be a nonempty bipartite graph on vertices a_1, a_2, b_1, b_2. Let (R, F) be a contributing fancy U-ribbon of length 2ℓ. Suppose that the sets*

$$\{F(a_i^1), F(b_i^1)\}, \quad \{F(a_i^2), F(b_i^2)\}, \quad \{F(c_i)\}$$

are disjoint. Then $\{F(u) : u \in R\}$ *contains at most* $3\ell + 2$ *distinct labels. If* U *is empty, then* $\{F(u) : u \in R\}$ *contains at most* $4\ell + 2$ *distinct labels.*

Proof. First suppose U is nonempty. By swapping a_i^1, a_i^2 or b_i^1, b_i^2 or both as necessary (which does not change whether (R, F) is contributing), we may assume that U contains the edge (a_1, b_1) and thus that R contains the edges (a_i^1, b_i^1) and (b_i^1, a_{i+1}^1) (where as usual addition is modulo ℓ).

Because (R, F) is contributing, every edge must have an identically-labeled partner. By our disjointness assumptions, edges among $\{a_i^1, b_i^1\}$ may be partnered only to edges similarly among $\{a_i^1, b_i^1\}$. Also, edges between $\{c_i\}$ and $\{a_i^2, b_i^2\}$ may be partnered only to edges between $\{c_i\}$ and $\{a_i^2, b_i^2\}$. Thus, the 2ℓ-edge-long cycle on vertices $\{a_i^1, b_i^1\}$ may have at most ℓ uniquely-labeled edges, and the 4ℓ-edge-long cycle on vertices $\{a_i^2, b_i^2, c_i\}$ may have at most 2ℓ uniquely-labeled edges. Since both are connected, the former may have at most $\ell + 1$ unique vertex labels and the latter at most $2\ell + 1$ unique vertex labels. Thus there are at most $3\ell + 2$ unique vertex labels in (R, F).

When U is empty the proof is similar: there are two paths, $\{a_i^1, b_i^1, c_i\}$ and $\{a_i^2, b_i^2, c_2\}$. □

4.4.4.3 Proofs of Lemma 4.4.11 and Lemma 4.4.12

Proof of Lemma 4.4.11. By Lemma 4.4.13 it is enough to prove the analogous claims for the $n^d \times n^d$ matrix Q with entries given by

$$Q[(a_1, \ldots, a_d), (b_1, \ldots, b_d)] = \begin{cases} \prod_{(i,j)\in B} g_{a_i, b_j} & \text{if } |\{a_1, \ldots, a_d, b_1, \ldots, b_d\}| = 2d \\ 0 & \text{otherwise} \end{cases},$$

By multiplying Q by suitable permutation matrices P, P' to give PQP', we may assume in the 2-matching case above that the matching is $\{(1,1),(2,2)\}$ and in the nonempty graph case that the edge contained is $(1,1)$ (where we think of the vertex set of U as $[d]\times[d]$). Note that $\|Q\| = \|PQP'\|$.

We apply Lemma 4.4.14 to obtain a family of matrices $\{Q_i\}_{i\in[r]}$ for some $r = O(3^3 \log n) = O(\log n)$ satisfying $Q = \sum_i Q_i$. On any entry $(a_1,\dots,a_d),(b_1,\dots,b_d)$ on which Q_i is nonzero it is equal to Q at that entry, and furthermore for each Q_i there is a partition $(S_1^i,\dots,S_3^i)$ of $[n]$ so that if $Q_i[(a_1,\dots,a_d),(b_1,\dots,b_d)] \neq 0$ then $a_1, b_1 \in S_1^i, a_2, b_2 \in S_2^i$, and $a_j, b_j \in S_3^i$ for all $j > 2$.

We show that every matrix $\|Q_i\|$ has bounded spectral norm. To save on indices, let $N = Q_i$. Let (S_1, S_2, S_3) be the partition of $[n]$ corresponding to N. We bound $\mathbb{E}\,\mathrm{Tr}(NN^\top)^\ell$ for some ℓ to be chosen later.

Let $\mathcal{R}(N)$ be the set of contributing disjoint labeled U-ribbons (R,F) of length 2ℓ with $F(a_1^i), F(b_1^i) \in S_1, F(a_2^i), F(b_2^i) \in S_2$ and $F(a_j^i), F(b_j^i) \in S_3$ for $j > 2$. Then $\mathbb{E}\,\mathrm{Tr}(NN^\top)^\ell \leqslant O(\ell^\ell)|\mathcal{R}(N)|$. (Here we have an inequality rather than an equality because some elements of $\mathcal{R}(N)$ may correspond to entries of N which are zero because they appeared in some other part of the partitioning scheme and $\ell^\ell \geqslant \ell!$ accounts for reorderings of the labels.)

Supposing that B contains a 2-matching, by Lemma 4.4.21, each $(R,F) \in \mathcal{R}(N)$ contains at most $(2d-2)\ell + 2$ unique $\{F(u) : u \in R\}$. So there are at most $n^{2\ell(d-1)+2}$ elements of $\mathcal{R}(N)$. It follows by Markov's inequality that for any $\alpha > 0$,

$$\mathbb{P}(\|N\| > \alpha) \leqslant \mathbb{P}(\mathrm{Tr}(NN^T)^\ell > \alpha^{2\ell}) \leqslant \frac{O(\ell^\ell) n^{2\ell(d-1)+2}}{\alpha^{2\ell}}$$

Choosing $\alpha \geqslant O(\ell) n^{d-1+10/\ell} (\log n)^{1/2\ell} 2^{d^2/2\ell}$ makes this at most $(\ell^\ell n^{10} \log(n) 2^{d^2})^{-1}$.

Choose $\ell = (\log n)^2$ so that there is such an α also satisfying $\alpha = O(n^{d-1}\log(n)^2)$ (so long as $d \leqslant O(\log n)$ as assumed).

Taking a union bound over the $\log n$ matrices Q_i, we get that

$$\mathbb{P}(\text{exists } i \text{ with} \|Q_i\| > O(\log(n)^2 n^{d-1})) \leqslant n^{-10}2^{-d^2}$$

and so by the triangle inequality applied to $\|M\| = \|\sum_i Q_i\|$, we get

$$\mathbb{P}(\|Q\| > O(n^{d-1}\log(n))^3) \leqslant n^{-10}2^{-d^2}\,.$$

The case that B contains only a 1-matching is similar, replacing the $(2d-2)\ell + 2$ unique vertices in a contributing B-ribbon with $(2d-1)\ell + 1$, again by Lemma 4.4.21. □

Proof of Lemma 4.4.12. We first handle the case when U is non empty. By Lemma 4.4.13 it is enough to prove the analogous statement for the $n^2 \times n^2$ matrix, also by abuse of notation denoted Q, which is the sum of matrices (abusing notation again) Q_k with entries given by

$$Q^k[(a_1,a_2),(b_1,b_2)] = \begin{cases} g_{k,a_1}g_{k,a_2}g_{k,b_1}g_{k,b_2}\prod_{(i,j)\in U} g_{a_i,b_j} & \text{if } |\{a_1,a_2,b_1,b_2\}| = 4 \\ 0 & \text{otherwise} \end{cases}.$$

By multiplication with an appropriate permutation matrix (which cannot change the spectral norm), we may assume that U contains the edge $(1,1)$. We begin with 2 from Lemma 4.4.14, whose hypotheses are satisfied by our convention $y_{a,a} = 0$. This gives a family $Q_1,\dots,Q_r$ with $r = O(3^3\log n) = O(\log n)$ so that $\sum_{i=1}^r Q_i$ and a corresponding family of partitions $(S_1^1,S_2^1,T^1),\dots,(S_1^r,S_2^r,T^r)$. Lemma 4.4.14 guarantees that $Q_i[(a_1,a_2),(b_1,b_2)] = \sum_{k\in T} y_{k,a_1}y_{k,a_2}y_{k,b_1}y_{k,b_2}\prod_{(i,j)\in U} y_{a_i,b_j}$ for some $T \subseteq T^i$ when $a_1,b_1 \in S_1^i$ and $a_2,b_2 \in S_2^i$ and is zero otherwise.

Fix some $i \in [r]$ and let $N = Q_i$ (to save on indices). We will bound $\mathbb{E}\operatorname{Tr}(NN^\top)^\ell$ for some ℓ to be chosen later. Let (S_1, S_2, T) be the partition of $[n]$ corresponding to N.

Let $\mathcal{R}(N)$ be the set of contributing labeled fancy U-ribbons of length 2ℓ so that for each $c_i \in R$, $F(c_i) \in T$, for each a_i^1, b_i^1 we have $F(a_i^1), F(b_i^1) \in S_1$, and for each a_i^2, b_i^2 we have $F(a_i^2), F(b_i^2) \in S_2$.

Expanding $\mathbb{E}\operatorname{Tr}(NN^\top)^\ell$ as usual, we see that $\mathbb{E}\operatorname{Tr}(NN^\top)^\ell \leqslant \ell^\ell |\mathcal{R}(N)|$. (As in the proof of Lemma 4.4.11, we have an inequality rather than an equality because some entries of N may not have a sum over all elements of T if there is overlap with previous parts of the partitioning scheme.) By Lemma 4.4.23, $|\mathcal{R}(N)| \leqslant \binom{n}{3\ell+2} \leqslant n^{3\ell+2}$.

By Markov's inequality,

$$\mathbb{P}(\|N\| > \alpha) \leqslant \mathbb{P}(\operatorname{Tr}(NN^\top)^\ell > \alpha^{2\ell}) \leqslant \frac{\ell^\ell n^{3\ell+2}}{\alpha^{2\ell}}.$$

Taking $\alpha \geqslant \ell n^{3/2+13\ell/2}$ guarantees that this is at most n^{-11}. If $\ell = \Theta(\log n)$, then there is such an α satisfying also $\alpha = O(n^{3/2}\log(n))$.

By a union bound and triangle inequality, we then get

$$\mathbb{P}(\|Q\| > O(n^{3/2}(\log n)^2) \leqslant O(n^{-10}). \qquad \square$$

The proof in the case of U empty is similar, using Lemma 4.4.23 in the empty U case.

4.5 Analyzing Deviations for the Degree-d MPW Operator

In this section, we use the tools developed in Section 4.4 to analyze the spectrum of the deviation matrix $D = \mathcal{M}' - E$ and prove Lemma 4.2.7.

We start by decomposing $\mathcal{M}' = E + D$. For any $I, J \in \binom{[n]}{d}$, $D(I,J)$ depends on a) $\deg(I \cup J)$ and b) whether $\mathcal{E}_{ext}(I,J) \subseteq G$. If $D(I,J)$ depended only on b) above, then it could be decomposed into a sum of patterned matrices defined in Section 4.4; analyzing these is tractable. Our first step is thus to get rid of the dependence on $\deg(I \cup J)$—the only part depending on the entire graph. We will obtain a matrix L that depends only on whether $\mathcal{E}_{ext}(I,J) \subseteq G$ or not (and thus is "locally random" in the sense of [85]).

Specifically, we write $D = L + \Delta$ where L is the locally random part obtained by replacing $D(I,J)$ by $\mathbb{E}[D(I,J) \mid \mathcal{E}_{ext}(I,J) \subseteq G]$ whenever $\mathcal{E}_{ext}(I,J) \subseteq G$ and an appropriate negative constant when $\mathcal{E}_{ext}(I,J) \not\subseteq G$ (this makes the expectation of each entry over $G \sim G(n, \frac{1}{2})$ to be 0). More concretely, following [85], we define:

Definition 4.5.1.

$$\alpha(i) = \frac{\binom{\omega}{2d-i}}{\binom{2d}{2d-i}} \cdot \binom{n-2d+i}{i} \cdot 2^{-d^2-\binom{i}{2}}, \tag{4.5.1}$$

and $p(i) = 2^{-(d-i)^2}$ for each i. We set $L(I,J)$ for every $I, J \in \binom{[n]}{d}$ to be

$$L(I,J) = \begin{cases} \alpha(|I \cap J|) \cdot \frac{(1-p(|I \cap J|))}{p(|I \cap J|)} \text{ if } \mathcal{E}(I \cup J) \setminus (\mathcal{E}(I) \cup \mathcal{E}(J)) \subseteq G \\ -\alpha(|I \cap J|), \text{ otherwise.} \end{cases}$$

We define $\Delta = D - L$.

The idea behind the definition is that $L(I,J) = \mathbb{E}_{G \sim G(n,\frac{1}{2})}[\mathcal{M}'(I,J) \mid \mathcal{E}_{int}(I,J) \subseteq G]$ whenever $\mathcal{E}_{ext}(I,J) \subseteq G$ and in the other case, chosen to make $\mathbb{E}_G[L(I,J)] = 0$. We will analyze L and Δ separately. The proof of Lemma 4.2.7 is broken into two main pieces. Each piece analyzes the action of L and Δ split across various eigenspaces $V_0, V_1, \ldots, V_d$ of the matrix E. Such fine grained analysis for the case of $d = 2$ was done in [43]. A few points of distinctions from [43] are in order at this point.

The first is regarding the high level approach. The approach of [43] used explicit expressions for a canonical set of eigenvectors in V_1 to obtain similar conclusions as us for the case of $d = 2$. This approach gets unwieldy very quickly because the explicit entries of eigenvectors for V_i for $i > 1$ are hard to work with [16]. We tackle this issue by developing an argument that doesn't need explicit entries of the eigenvectors. Instead, we use basic representation theory (Section 4.4.3) to identify a set of symmetries satisfies by vectors in V_i for each i and use it obtain the conclusions we require.

Second, [43] deal with the optimization version of the degree 4 SOS program which, as noted in the introduction, could be potentially weaker than the one we analyze here (and thus our lower bound is technically stronger). This simplifies the analysis in [43] a little bit as the matrix Δ defined above is identically zero for the operator analyzed. We explicitly work with the feasibility version of the degree 4 SOS program and thus, must deal with the additional complexity of handling Δ. It turns out that we have to do a fine grained analysis of the Δ matrix itself. The decomposition we use for Δ is somewhat different from the case of L even though, the analysis of each piece of the decomposition proceeds similar to the case of L.

Third, for the special case of $d = 2$, essentially the only matrix one has to analyze is the L_0, the matrix obtained by zeroing out all entries (I, J) in L such that $I \cap J \neq \emptyset$: a uniform bound on spectral norm of the remaining component suffices. However, for higher d, one has to deal with the "non-disjoint" entries with some care and an argument analogous to the one in [43] fails to show PSDness of $\mathcal{M}'$ beyond $\omega \approx n^{\frac{1}{2d}}$ giving no asymptotic improvement over [85].

Finally, our argument for analyzing the spectral norms of each of the pieces also

needs to be much more general than in case of [43] to handle higher degrees. For this, we identify a simple combinatorial structure (size of maximum matchings in appropriate bipartite graphs on $2d$ vertices) that controls the bounds and could also be used to obtain slick proofs of the conclusions required in [43] in the context of analyzing $\mathcal{M}'$ for $d = 2$. Our combinatorial argument itself is a generalization of the one given by [43] for this case.

We now go on to describe the two lemmas that encapsulate the technical heart of the proof of Lemma 4.2.7. The first does a fine grained analysis of spectrum of L. In the following, let $\mathcal{M}'$ be the filled-in matrix for the degree-d MPW operator at clique size ω (Definition 4.2.5), $E = \mathbb{E}_{G\sim G(n,1/2)}[\mathcal{M}']$, $D = \mathcal{M}' - E$, L be as Definition 4.5.1, Π_i be the projectors to the spaces V_i of Fact 4.2.9.

Lemma 4.5.2 (Bounding Blocks of L)**.** *With probability at least $1 - \frac{1}{n}$ over the draw of $G \sim G(n, \frac{1}{2})$, each $0 \leqslant i \leqslant d$ satisfies*

1. $$\left|\Pi_i L \Pi_j\right| \leqslant 2^{2d}\tilde{O}(\omega^{2d} n^{d-\frac{1}{2}}),$$

2. *If $i, j \geqslant 2$, then*
$$\left|\Pi_i L \Pi_j\right| \leqslant 2^{2d}\tilde{O}(\omega^{2d} n^{d-1}).$$

The next lemma does a (even more) fine grained analysis of the spectrum of Δ:

Lemma 4.5.3. *With probability at least $1 - \frac{1}{n}$ over the draw of $G \sim G(n, \frac{1}{2})$, for each $0 \leqslant i \leqslant d$:*

$$\left|\Pi_i \Delta \Pi_j\right| \leqslant \tilde{O}(2^{O(d)}\omega^{2d-\min\{i,j\}} n^{d-\frac{1}{2}}) + \tilde{O}(2^{O(d)}\omega^{2d-q} n^{d-1}).$$

We can now use Lemma 4.5.2 and Lemma 4.5.3 to complete the proof of Lemma 4.2.7.

Proof of Lemma 4.2.7. For each $0 \leqslant i, j \leqslant d$, we compute $\Pi_i \mathcal{M}' \Pi_j$ and use Lemma 4.1.2. We write $\mathcal{M}' = E + L + \Delta$. First, $\Pi_i^\dagger E \Pi_j = 0$ whenever $i \neq j$ as Π_i are projectors to eigenspaces of E. Let $\lambda_0, \lambda_1, \ldots, \lambda_d$ be the eigenvalues of E on eigenspaces $V_0, V_1, \ldots, V_d$. From Lemma 4.2.10, we have:

$$\lambda_j \geqslant 2^{-O(d^2)} \cdot n^d \cdot \omega^{2d-j}.$$

Thus,

$$\Pi_i E \Pi_i \geqslant 2^{-O(d^2)} n^d \omega^{2d-j}.$$

In what follows, all our statements hold with probability at least $1 - O(1)/n$:

Using Lemma 4.5.2 and Lemma 4.5.3, for every $i \geqslant 2$,

$$\|\Pi_i (L + \Delta) \Pi_i\| \leqslant \tilde{O}(\omega^{2d} n^{d-1}) + \tilde{O}(\omega^{2d-i} n^{d-\frac{1}{2}}).$$

On the other hand, when $i \leqslant 2$,

$$\|\Pi_i (L + \Delta) \Pi_i\| \leqslant \tilde{O}(\omega^{2d} n^{d-\frac{1}{2}}).$$

Then, it is easy to check that for any $\omega = O(n^{\frac{1}{d+1}})$,

$$\Pi_i \mathcal{M}' \Pi_i = \Pi_i (E + L + \Delta) \Pi_i \geqslant 2^{-O(d^2)} \omega^{2d-i} n^d.$$

Next, we bound the cross terms $|\Pi_i (L + \Delta) \Pi_j|$ for $i \neq j$. Again, using Lemma 4.5.2 and Lemma 4.5.3, we have for $i, j \geqslant 2$:

$$|\Pi_i(L+\Delta)\Pi_j| \leqslant \tilde{O}(\omega^{2d} n^{d-1}) + \tilde{O}(\omega^{2d-\min\{i,j\}} n^{d-\frac{1}{2}}).$$

For $\omega \leqslant O(n^{\frac{1}{d+1}})$, it is again easy to check that, for $i, j \geqslant 2$, the above expression is at most $\frac{2}{d}\sqrt{\lambda_i \lambda_j}$.

In the case when one of i, j is at most 1, we have the bound

$$|\Pi_i(L+\Delta)\Pi_j| \leqslant \tilde{O}(\omega^{2d} n^{d-\frac{1}{2}}).$$

In this case, it is easy to check that so long as $\omega = O(n^{\frac{1}{d+1}} / \operatorname{poly}\log(n))$,

$$|\Pi_i(L+\Delta)\Pi_j| \leqslant \tilde{O}(\omega^{2d} n^{d-\frac{1}{2}}) \leqslant \frac{2}{d}\sqrt{\lambda_i \lambda_j}.$$

By an application of Lemma 4.1.2, the proof is complete. □

4.5.1 Proof of Lemma 4.5.2

Proof Plan. We first describe the high level idea of the proof.

We start by decomposing $L = \sum_{q=0}^{d} L_q$ where $L_q(I, J) = L(I, J)$ if $|I \cap J| = q$ and 0 otherwise. Notice that each L_q then is obtained by a scaling an appropriate 0/1 matrix.

Most illuminating is the *disjoint* case L_0, which is nonzero only at entries I, J with $I \cap J = \emptyset$. For any disjoint I, J, $L_0(I, J)$ depends only whether $\mathcal{E}_{ext}(I, J) \subseteq G$, which, one could write as an appropriately scaled AND function of the indicators g_e of edges $e \in \mathcal{E}_{ext}(I, J)$. We can expand this AND function in the monomial (parities of subsets of g_e variables) basis. Each such monomial corresponds to the bipartite graph B that contains the pairs $e \in \mathcal{E}_{ext}(I, J)$ that constitute the monomial. This gives a decomposition of L_0

into $2^{d^2} - 1$ (since the constant term is 0, L being zero mean) components, L_0^B for each non empty, labeled bipartite graph on $[d] \times [d]$.

We can bound the spectral norm of each of the pieces L_0^B by direct application of tools derived in Section 4.4.4. The main work in this section goes into showing that depending upon the structure of B, an appropriate selection of subspaces V_i lie in left or right kernels of L_0^B. Thus, for a fixed term $\Pi_i L \Pi_j$, some L_0^B do not contribute. We identify the maximum spectral norm among contributing terms to obtain the final bound.

To accomplish this goal, we rely heavily on the tools built in Section 4.4.3 which give us a handle on the symmetries of the eigenspaces $V_0, V_1, \ldots, V_d$. This requires some work based on representation theory of finite groups and is presented in Lemma 4.4.4 and Lemma 4.4.9.

The case of L_q for $q \neq 0$ needs even finer decomposition. We decompose each L_q ($q > 0$) into matrices that identify the "pattern" of the q intersecting vertices. In [85] a similar idea is used to reduce the task of bounding the spectral norm of L_q to a calculation similar to one in the case of L_0. However, unlike [85], we also require properties of the kernels of the components of the decomposition. After restricting to a fixed intersection pattern of q vertices, we thus resort to using a generalization of the kernel analysis used for the L_0 case. We now proceed with the proof plan as described beginning with the decomposition of each L_q.

4.5.1.1 Decomposing L

We start by decomposing L further as $L = \sum_{q=0}^{d} L_q$ where for any $I, J \in \binom{[n]}{d}$,

$$L_q = \begin{cases} L(I,J) & \text{if } |I \cap J| = q, \\ 0 & \text{otherwise .} \end{cases}$$

Decomposing L_0. Recall that $\mathcal{B}$ is the set of all bipartite labeled graphs with left and right vertex sets labeled by $[d]$ and $[d]$. Recall also from Section 4.4.3 that for any $I, J \subseteq [n]$ with $|I| = |J| = d$, the graph $\zeta_{I,J}(B)$ is a copy of B on vertex sets I, J where the correspondence between I and $[d]$ and J and $[d]$ is determined by the sorting map ζ. Finally, recall that for a graph G on $[n]$, we let g_b be the ± 1 indicator for the presence of edge b in G, and by convention $g_b = 0$ when $b = (i,i)$ for any $i \in [n]$. For any $B \in \mathcal{B}$, define an $\binom{[n]}{d} \times \binom{[n]}{d}$ matrix

$$\tilde{L}_0^B(I,J) = \alpha(0) \cdot (2^{d^2} - 1)\Pi_{b \in \zeta_{I,J}(B)} g_b.$$

The idea is to write L_0 as a sum of such matrices $\tilde{L}_0^B$ with the entries corresponding I, J where $|I \cap J| \neq 0$ zeroed out. Thus, define L_0^B to be the matrix with (I,J) entry given by

$$L_0^B(I,J) = \begin{cases} \tilde{L}_0^B(I,J) & \text{if } |I \cap J| = 0 \\ 0 & \text{otherwise.} \end{cases}$$

We think of L_0 as a rescaling and centering of a 0/1 matrix whose entries are the AND of the ± 1 indicators for the edges in $\mathcal{E}_{ext}(I,J)$. Decomposing these ANDs into monomials over those ± 1 indicators, we see that each monomial corresponds exactly to one bipartite graph B, and the centering of L_0 corresponds to removing the constant monomial, which corresponds to the empty bipartite graph. Every other monomial recieves equal weight

2^{-d^2} in this expansion, and so from these observations it becomes routine to verify that

$$2^{-d^2} \sum_{B \in \mathcal{B}} L_0^B = L_0.$$

Decomposing L_q. Similarly, we further decompose L_q for $q > 0$. Here things are a bit more involved. Let us motivate our decomposition by understanding the structure of the matrix L_q for $q > 0$ a little bit. Consider an entry (I, J) such that $I \cap J = K$. Then, $\mathcal{E}_{ext}(I, J) \subseteq \mathcal{E}_{ext}(I \setminus K, J \setminus K)$. Thus, the edge structure in the bipartite subgraph on vertex sets $I \setminus K$ and $J \setminus K$ decides the value of $L(I, J)$ for any graph G and we can hope to a get a patterned matrix. We now follow this intuition.

Recall the sorting maps $\zeta_I : [d] \to I$ and $\zeta_J : [d] \to J$. Letting $Z_\ell, Z_r \subseteq [d]$ be subsets of size q, we define $L_q^{Z_\ell, Z_r}$ such that:

$$L_q^{Z_\ell, Z_r}(I, J) = \begin{cases} L_q(I, J) & \text{if } \zeta_I(Z_\ell) = \zeta_J(Z_r) \\ 0 & \text{otherwise.} \end{cases}$$

That is, $L_q^{Z_\ell, Z_r}$ is the "part" of L_q where any I, J intersect in a (size q) subset given by $\zeta_I(Z_\ell)$ and $\zeta_J(Z_r)$. It is then easy to see that $L_q = \sum_{Z_\ell, Z_r} L_q^{Z_\ell, Z_r}$. Next, we decompose each $L_q^{Z_\ell, Z_r}$ further based on non-empty labeled bipartite graphs $B \in \mathcal{B}_{Z_\ell, Z_r}$ for each Z_ℓ, Z_r.

We now define a matrix which is nonzero only on entries which intersect in *at least* q places:

$$\tilde{L}_q^{B, Z_\ell, Z_r}(I, J) = \begin{cases} \alpha(q) \cdot (2^{(d-q)^2} - 1) \Pi_{b \in \zeta_{I,J}(B)} g_b & \text{if } \zeta_I(Z_\ell) = \zeta_J(Z_r) \\ 0 & \text{otherwise.} \end{cases}$$

Again, as before, the actual decomposition needs to zero out the entries (I, J) such that $|I \cap J| \neq q$. Thus, we define L_q^{B, Z_ℓ, Z_r} by zeroing entries of $\tilde{L}_q^{B, Z_\ell, Z_r}$ which intersect also

outside of Z_ℓ, Z_r:

$$L_q^{B,Z_\ell,Z_r}(I,J) = \begin{cases} \tilde{L}_q^{B,Z_\ell,Z_r}, \text{ if} |\zeta_I([d] \setminus Z_\ell) \cap \zeta_J([d] \setminus Z_r)| = 0 \\ 0, \text{ otherwise.} \end{cases}$$

Finally, it is again easy to verify that:

$$L_q = 2^{-(d-q)^2} \sum_{Z_\ell, Z_r \subseteq [d]} \sum_{B \in \mathcal{B}_{Z_\ell,Z_r}} L_q^{B,Z_\ell,Z_r}.$$

4.5.1.2 Spectral Analysis of L

In order to prove Lemma 4.5.2, we will first use the decomposition described in the previous section to write L_q as a sum of appropriate patterned matrices. We will then partition the sum into groups, each group corresponding to an equivalence class of (left or right) similar bipartite graphs B. We will infer some properties about the kernel and finally use the spectral norm bounds from Section 4.4.4 to complete the proof.

More concretely, let (B, Z_ℓ, Z_r) be a q pattern (as defined in Section 4.4.3). From our decomposition from the previous section, we have:

$$L_q = 2^{-(d-q)^2} \sum_{Z_\ell, Z_r} \sum_{B \in \mathcal{B}_{Z_\ell,Z_r}} L_q^{B,Z_\ell,Z_r} \tag{4.5.2}$$

Our idea is to analyze appropriate collections of L_q^{B,Z_ℓ,Z_r} separately. When $q = 0$, Z_ℓ, Z_r are redundant (being $\emptyset$) and thus $L_0^{B,Z_\ell,Z_r} = L_0^B$ in that special case. In the first step, we observe that $\tilde{L}_q^{B,Z_\ell,Z_r}$ has some symmetries that are helpful to us. We thus want to deal with the sums of $\tilde{L}_q^{B,Z_\ell,Z_r}$ instead of L_q^{B,Z_ℓ,Z_r}. To justify this, we start by showing that the difference of the above two matrices has small norm.

Claim 4.5.4. For any (B, Z_ℓ, Z_r),

$$||L_q^{B,Z_\ell,Z_r} - \tilde{L}_q^{B,Z_\ell,Z_r}|| \leqslant \tilde{O}(\omega^{2d-q} n^{d-1}).$$

Next, we bound each $\|\tilde{L}_q^{B,Z_\ell,Z_r}\|$ using the machinery from Section 4.4.4.

Claim 4.5.5 (Norm Bounds on Pieces). With probability at least $1 - 1/n^{10}$ over the draw of $G \sim G(n, \frac{1}{2})$,

1.
$$\|\tilde{L}_q^{B,Z_\ell,Z_r}\| = \tilde{O}(\omega^{2d-q} \cdot n^{d-\frac{1}{2}}).$$

2. If $|B_\ell|, |B_r| \geqslant 2$, then:
$$\|\tilde{L}_q^{B,Z_\ell,Z_r}\| = \tilde{O}(\omega^{2d-q} \cdot n^{d-1}).$$

At first, it should be worrisome that some of the $\tilde{L}_q^{B,Z_\ell,Z_r}$ have norms that are much larger than what we need (in the second claim of Lemma 4.5.2). What comes to our rescue is the fact that the components $\tilde{L}_q^{B,Z_\ell,Z_r}$ that have large norm do not contribute to quadratic forms on $\left|\Pi_i^\dagger L \Pi_j\right|$ when i, j are at least 2. The crucial observation that allows us to conclude this is based on the observation that $\tilde{L}_q^{B,Z_\ell,Z_r}$ are patterned matrices in the sense of Definition 4.4.8 and thus clubbing all (B', Z'_ℓ, Z'_r) that are (left or right) similar to (B, Z_ℓ, Z_r), we can show that certain V_i lie in their kernels. More specifically:

Claim 4.5.6. For $t > q + |B_\ell|$,

$$\Pi_t^\dagger \left(\sum_{(B,Z'_\ell,Z'_r) \sim_\ell (B,Z_\ell,Z_r)} \tilde{L}_q^{B',Z'_\ell,Z'_r} \right) = 0.$$

Similarly, for any $w > q + |B_r|$,

$$\left(\sum_{(B',Z'_\ell,Z'_r)\sim_r(B,Z_\ell,Z_r)} \tilde{L}_q^{B,Z_\ell,Z_r}\right)\Pi_w = 0.$$

Before proving the three claims above, we show how they imply Lemma 4.5.2:

Proof. We use (4.5.2) to write:

$$L_q = 2^{-(d-q)^2}\left(\sum_{B,Z_\ell,Z_r} \tilde{L}_q^{B,Z_\ell,Z_r} + \sum_{B,Z_\ell,Z_r} (L_q^{B,Z_\ell,Z_r} - \tilde{L}_q^{B,Z_\ell,Z_r})\right) \tag{4.5.3}$$

$$\begin{aligned}
L_q &= 2^{-(d-q)^2}\left(\sum_{B,Z_\ell,Z_r} \tilde{L}_q^{B,Z_\ell,Z_r} + \sum_{B,Z_\ell,Z_r} (L_q^{B,Z_\ell,Z_r} - \tilde{L}_q^{B,Z_\ell,Z_r})\right) \\
&\leqslant 2^{-(d-q)^2}\left(\sum_{B,Z_\ell,Z_r} \|\tilde{L}_q^{B,Z_\ell,Z_r}\|\right) + 2^{-(d-q)^2}\cdot \sum_{B,Z_\ell,Z_r} \|L_q^{B,Z_\ell,Z_r} - \tilde{L}_q^{B,Z_\ell,Z_r}\| \\
&\quad \text{Using Claim 4.5.5} \\
&\leqslant 2^{2d}\cdot \tilde{O}(\omega^{2d-q} n^{d-\frac{1}{2}}).
\end{aligned}$$

For the second part, fix an $i \geqslant 2$. We first show that some terms in the decomposition in (4.5.3) do not contribute to $|\Pi_i^\dagger L_q \Pi_i|$.

Consider any bipartite graph B such that $|B_\ell| < 2$. Then, we have from Claim 4.5.6, $\left(\sum_{(B',Z'_\ell,Z'_r)\sim_\ell(B,Z_\ell,Z_r)} L_q^{B',Z'_\ell,Z'_r}\right)\Pi_t = 0$ for every $t \geqslant 2$. Thus, $|\Pi_i^\dagger \sum_{B:|B_\ell|<2} L_q^{B,Z_\ell,Z_r}\Pi_i| = 0$ for any i and every $t \geqslant 2$. Similarly, when $|B_r| \geqslant 2$, $|\Pi_i^\dagger L_q^{B,Z_\ell,Z_r}\Pi_i| = 0$. On the other hand, when both $|B_\ell|, |B_r| \geqslant 2$, from Claim 4.5.5, we have (with high probability over the draw of $G \sim G(n, \frac{1}{2})$), $\|L_q^{B,Z_\ell,Z_r}\| \tilde{O}(\omega^{2d-q}\cdot n^{d-1})$. Thus:

Thus, for $i \geqslant 2$, we have:

$$\begin{aligned}
||\Pi_i^\dagger L_q \Pi_i|| &= 2^{-(d-q)^2} \cdot \| \sum_{B,Z_\ell,Z_r} \Pi_i^\dagger L_q^{B,Z_\ell,Z_r} \Pi_i \| \\
&\leqslant 2^{-(d-q)^2} \sum_{B,Z_\ell,Z_r} |\Pi_i^\dagger \tilde{L}_q^{B,Z_\ell,Z_r} \Pi_i| + 2^{-(d-q)^2} \cdot \sum_{B,Z_\ell,Z_r} |L_q^{B,Z_\ell,Z_r} - \tilde{L}_q^{B,Z_\ell,Z_r}| \\
&= 2^{-(d-q)^2} \sum_{B:|B_\ell|\geqslant 2,|B_r|\geqslant 2,Z_\ell,Z_r} |\Pi_i^\dagger \tilde{L}_q^{B,Z_\ell,Z_r} \Pi_i| + \tilde{O}(\omega^{2d} \cdot n^{d-1}) \\
&\leqslant \tilde{O}(\omega^{2d-q} \cdot n^{d-\frac{1}{2}}).
\end{aligned}$$

Similarly, for $i + j \geqslant 2$ we must have $i, j \geqslant 2$. Thus, by a calculation similar to above, $\|\Pi_i^\dagger L_q \Pi_j\| \leqslant \tilde{O}(\omega^{2d} \cdot n^{d-\frac{1}{2}})$. □

In the remaining part of this section, we complete the proofs of Claims 4.5.6 and 4.5.5.

4.5.1.3 Proof of Claims

In this section, we obtain quick proofs of the three claims above using the tools developed in Section 4.4.

We first prove Claim 4.5.4.

Proof of Claim 4.5.4. The proof is by appealing to Fact 4.1.1. Observe that

$$|\tilde{L}_q^{B,Z_\ell,Z_r}(I,J) - L_q^{B,Z_\ell,Z_r}(I,J)| \leqslant \begin{cases} 0 \text{ , if } |I \cap J| \leqslant q, \\ \alpha(q) \cdot (2^{(d-q)^2} - 1) 2^{-(d-q)^2} \text{ , otherwise.} \end{cases}$$

We now estimate

$$\max_{I \in \binom{[n]}{d}} \sum_{J \in \binom{[n]}{d}} |\tilde{L}_q^{B,Z_\ell,Z_r}(I,J) - L_q^{B,Z_\ell,Z_r}(I,J)| \leqslant 2^d n^{d-q-1} \cdot \alpha(q) \cdot (2^{(d-q)^2} - 1) 2^{-(d-q)^2}.$$

The claim now follows from Fact 4.1.1. □

The next is a direct application of Lemma 4.4.9.

Proof of Claim 4.5.6. The main observation is that $\tilde{L}_q^{B,Z_\ell,Z_r}$ is a patterned matrix (with a q-pattern (B, Z_ℓ, Z_r)) in the sense of Definition 4.4.8. The result then follows immediately by appealing to Lemma 4.4.9. □

Finally, we prove Claim 4.5.5 using Lemma 4.4.11.

Proof of Claim 4.5.5. First, consider the case of $\tilde{L}_0^B$ ($Z_\ell = Z_r = \emptyset$ in this case). We write $||\tilde{L}_0^B|| \leqslant ||L_0^B|| + ||\tilde{L}_0^B - L_0^B||$. For the second term, we can appeal to Claim 4.5.4. For the first term, observe that by a direct application of Lemma 4.4.11, $||L_0^B|| \leqslant \alpha(0)(2^{d^2} - 1)2^{-d^2} \cdot \tilde{O}(n^{d-\frac{1}{2}})$. Further, when $|B_\ell|, |B_r| \geqslant 2$, then, B has a 2-matching and thus, by another application of Lemma 4.4.11, we obtain that in this case, $||L_0^B|| \leqslant \alpha(0) \cdot (2^{d^2} - 1)2^{-d^2} \cdot \tilde{O}(n^{d-1})$. The proof is thus complete for the case of $q = 0$.

We now reduce the computation for the more general case to similar calculations by appealing to the idea of lifts. Consider the matrix $R \in \mathbb{R}^{\binom{[n]}{d-q} \times \binom{[n]}{d-q}}$ given by

$$R(I', J') = \begin{cases} L_q^{B,Z_\ell,Z_r}(I' \cup K, J' \cup K) \text{ if } I' \cap J' = \emptyset. \\ 0, \text{ otherwise,} \end{cases}$$

where $K \subseteq [n]$ is some fixed subset of size q such that $K \cap I' = K \cap J' = \emptyset$.

Then, $L_q^{B,Z_\ell,Z_r} = R^{(q)}$. Thus, using Fact 4.4.16, $\|L_q^{B,Z_\ell,Z_r}\| \leqslant 2^{2d}\|R\|$. Since R has non zero entries only when the row and column indices are disjoint sets, we can apply Lemma 4.4.11 to R to obtain $\|R\| \leqslant \alpha(q)(2^{(d-q)^2} - 1)2^{-(d-q)^2} \cdot \tilde{O}(n^{d-q-\frac{1}{2}})$. Further, when

$|B_\ell|, |B_r| \geqslant 2$, again by an application of Lemma 4.4.11, we have:$\|R\| \leqslant \alpha(q)(2^{(d-q)^2} - 1)2^{-(d-q)^2} \cdot \tilde{O}(n^{d-q-1})$. Now, using $\|\tilde{L}_q^{B,Z_\ell,Z_r}\| \leqslant \|L_q^{B,Z_\ell,Z_r}\| + \|\tilde{L}_q^{B,Z_\ell,Z_r} - L_q^{B,Z_\ell,Z_r}\|$ and using Claim 4.5.4 completes the proof. □

4.5.2 Proof of Lemma 4.5.3

We now move on to analyzing the spectrum of the matrix Δ.

The high level plan of the proof is similar to that of Lemma 4.5.2. We define Δ_i for each $0 \leqslant i \leqslant d$ as follows:

$$\Delta_q(I, J) = \begin{cases} \Delta(I, J), \text{ if } |I \cap J| = q \\ 0, \text{ otherwise.} \end{cases}$$

We further split $\Delta_q = \sum_{K:|K|=q} \Delta_{q,K}$, where:

$$\Delta_{q,K}(I, J) = \begin{cases} \Delta_q(I, J), \text{ if} I \cap J = K \\ 0, \text{ otherwise.} \end{cases}$$

First, we observe that $\Delta_0 = 0$. This is because $\deg(I \cup J)$ when I and J are disjoint is exactly 1 and doesn't depend on the graph. Thus, $\Delta = \sum_{i=1}^{n} \Delta_i$. As before we would like to spot patterned matrices in each Δ_i to show that appropriate eigenspaces V_j lie in the kernel of Δ_i. In case of Δ_q, however, there's a difference how this needs to be done. This is because each entry (I, J) of Δ_i potentially depends on the edges from every vertex in the graph G to I and J. This is unlike the case of L where the (I, J) entry depends only on the edges between I and J (in fact that's the reason we separated L from Δ in the analysis). Nevertheless, we give a decomposition below that will help us make claims similar to the ones in the case of analyzing L in this case too.

Let us first explain the main idea in the decomposition. The entry $\Delta_q(I, J)$ depends on two events: a) whether $\mathcal{E}_{ext}(I, J) \subseteq G$ and b) the number of subsets $S \subseteq [n]$ of size $|S| = q$

that form $2d$-cliques with $I \cup J$. The main observation that motivates our decomposition is the following: in the event that $\mathcal{E}_{ext}(I, J) \subseteq G$, the deviation in $\deg(I \cup J)$ is completely captured (up to low order terms) by just the number of vertices s that has an edge to all of $I \cup J$ in G. This allows us to write entries of Δ_q as a sum of contribution to the deviation due to each vertex s separately. For the case of $q = 1$, this argument is in fact exact and there are no low order terms. When $q > 1$, the contributions due to individual vertices contribute the bulk of the deviation and only low order terms remain.

From here on, we are in a situation similar to the one encountered in analyzing L in the previous subsection. We show that the components in the decomposition with large spectral norm do not contribute to quadratic forms over eigenspaces with small eigenvalues of the expectation matrix E using the idea of patterned matrices from Section 4.4.3. We show that the remaining components have small spectral norm using the combinatorial techniques combined with the trace moment method developed in Section 4.4.4.

We now proceed to make the ideas above more precise. We begin by some notation and a definition. We first define $e^1_{s,K} \in \mathbb{R}^{\binom{[n]}{d}}$ for any $s \in [n]$ and $K \subseteq [n]$, $|K| = q$ as follows:

$$e^1_{s,K}(I, J) = \begin{cases} 0 & \text{, if } I \cap J \neq K \text{ or } s \in I \cup J \\ 2^{|I|-q} - 1 & \text{otherwise and if } \mathcal{E}_{ext}(\{s\}, I \setminus J) \subseteq G \\ -1, & \text{otherwise.} \end{cases}$$

Similarly, we define $e^2_s \in \mathbb{R}^{\binom{[n]}{d}}$ for any $s \in [n]$ and $K \subseteq [n]$, $|K| = q$:

$$e^2_{s,K}(I, J) = \begin{cases} 0 & \text{, if } I \cap J \neq K \text{ or } s \in I \cup J \\ 2^{|J|-q} - 1 & \text{otherwise and if } \mathcal{E}_{ext}(\{s\}, J \setminus I) \subseteq G \\ -1, & \text{otherwise.} \end{cases}$$

Next, we define: $e_{s,K} \in \mathbb{R}^{\binom{[n]}{d}}$ for any $s \in [n]$ and $K \subseteq [n]$ satisfying $|K| = q$:

$$e^3_{s,K}(I,J) = \begin{cases} 0 & \text{, if } I \cap J \neq K \text{ or } s \in I \cup J \\ 2^q - 1 \text{ otherwise and if } \mathcal{E}_{ext}(\{s\}, K) \subseteq G & \\ -1, \text{ otherwise.} & \end{cases}$$

Finally, we define

$$e^{1,2}_K(I,J) = \begin{cases} 0 & \text{, if } I \cap J \neq K \\ 2^{|I|+|J|-q} - 1 \text{ otherwise and if } \mathcal{E}_{ext}(I,J) \subseteq G & \\ -1, \text{ otherwise.} & \end{cases}$$

Using the matrices above, we can show the following approximate factorization for the entries of $\Delta_{q,K}$:

Lemma 4.5.7. *For every I, J such that $|I \cap J| = K$,*

$$\left| \Delta_{q,K}(I,J) - \eta \frac{\binom{\omega}{2d-q}}{\binom{2d}{2d-q}} (1 + e^{1,2}_K(I,J)) \sum_{s \in [n]} \left((1 + e^1_{s,K})(1 + e^2_{s,K})(1 + e^3_{s,K}) - 1 \right) \right| \leqslant 2^{O(d)} \cdot \tilde{O}(\omega^{2d-q} \cdot n^{q-1}),$$

for $\eta = 2^{-2|I \cup J| + \binom{|I \cup J|+1}{2} - \binom{2d}{2}} \frac{(n-|I|)^{2d-|I|-1}}{(2d-|I|-1)!}$.

Proof of Lemma 4.5.7. Let $A_{I \cup J}$ be the set of vertices s in G not in $I \cup J$ so that $(s,i) \in G$ for all $i \in I \cup J$. By definition, if $I \cup J$ is a clique,

$$\sum_{s \in [n]} (1 + e^{1,2}_{s,K}(I,J)) \left((1 + e^1_{s,K})(1 + e^2_{s,K})(1 + e^3_{s,K}) - 1 \right) = 2^{2|I \cup J|} (|A_I| - 2^{-|I \cup J|}(n - |I \cup J|)) .$$

Applying the scaling η, we get

$$\eta(1 + e^{1,2}(I,J)) \sum_{s \in [n]} \left((1 + e^1_{s,K})(1 + e^2_{s,K})(1 + e^3_{s,K}) - 1 \right) = \left(|A_I| - \frac{n - |I|}{2^{|I|}} \right) \frac{2^{\binom{|I|+1}{2} - \binom{2d}{2}} (n - |I|)^{2d-|I|-1}}{(2d - |I| - 1)!}$$

and hence the lemma follows from Theorem 4.7.4. □

We will need another definition before proceeding: For each i,

$$e^i_{s,q} \stackrel{\text{def}}{=} \sum_{K:|K|=q} e^i_{s,K}$$

As in the case of analyzing L, we define the filled in versions of $e^i_{s,K}$ as follows:

$$\tilde{e}^1_{s,K}(I,J) \stackrel{\text{def}}{=} \begin{cases} 0 & \text{, if } K \not\subseteq I \cap J \text{ or } s \in I \cup J \\ 2^{|I|-q} - 1 \text{ otherwise and if } \mathcal{E}_{ext}(\{s\}, I) \subseteq G & \\ -1, \text{ otherwise.} & \end{cases}$$

$$\tilde{e}^2_{s,K}(I,J) \stackrel{\text{def}}{=} \begin{cases} 0 & \text{, if } K \not\subseteq I \cap J \text{ or } s \in I \cup J \\ 2^{|J|-q} - 1 \text{ otherwise and if } \mathcal{E}_{ext}(\{s\}, J) \subseteq G & \\ -1, \text{ otherwise.} & \end{cases}$$

$$\tilde{e}^1_{s,K}(I,J) \stackrel{\text{def}}{=} \begin{cases} 0 & \text{, if } K \not\subseteq I \cap J \text{ or } s \in I \cup J \\ 2^{q} - 1 \text{ otherwise and if } \mathcal{E}_{ext}(\{s\}, K) \subseteq G & \\ -1, \text{ otherwise.} & \end{cases}$$

We start by giving norm bounds on all the matrices involved in the decomposition in Lemma 4.5.7.

Lemma 4.5.8. *1. For each $i \in \{1,2,3\}$,*

$$\left\| \sum_{s\in[n],K:|K|=q} e^i_{s,K} \right\| \leqslant \tilde{O}(2^{O(d)} \cdot n^{d-\frac{1}{2}}).$$

2. For each $i \in \{1,2,3\}$,

$$\left\| \sum_{s\in[n],K:|K|=q} e^i_{s,K} - \tilde{e}^i_{s,K} \right\| \leqslant \tilde{O}(2^{O(d)} \cdot n^{d-3/2}).$$

3.

$$\left\| \sum_{K:|K|=q} (1 + e^{1,2}_K(I,J)) \odot \sum_{s\in[n]} \left((1 + e^1_{s,K}) \odot (1 + e^2_{s,K}) \odot (1 + e^3_{s,K}) - \sum_{i=1}^{3} e^i_{s,K} \right) \right\| \leqslant \tilde{O}(2^{O(d)} \cdot n^{d-1}).$$

Proof. Fix a q and let Q be any matrix in $\mathbb{R}^{\binom{[n]}{d}\times\binom{[n]}{d}}$ that appears in the statement of the lemma above. Let $R \in \mathbb{R}^{\binom{[n]}{d-q}\times\binom{[n]}{d-q}}$ be defined by

$$R(I,J) = \begin{cases} \sum_{s\in[n],K:|K|=q} Q(I\cup K, J\cup K), \text{ if} I\cap J = \emptyset \\ 0\text{ , otherwise.} \end{cases}$$

Then, $\sum_{s\in[n],K:|K|=q} Q = R^{(q)}$ in the sense of Definition 4.4.15. Thus, by Fact 4.4.16 $\|R^{(q)}\| \leqslant 2^{2d}\|R\|$. Thus, we focus on bounding $\|R\|$ in the following. We will use the trace moment method for this purpose and the argument is similar to the ones made in Section 4.4.4 to develop the general purpose spectral concentration results. For this reason, we will be a bit more terse than in the case of the other applications of the trace moment method before. We set up the notation for the general R as above and specialize the combinatorial reasoning for each of the specific matrices involved later.

We expand

$$\mathbb{E}[\mathrm{Tr}((RR^\dagger)^\ell] = \mathbb{E}[\sum_{K_1,K_2,\ldots,K_{2\ell}} \sum_{I_1,I_2,\ldots,I_{2\ell},s_1,s_2,\ldots,s_{2\ell}} R(I_1\setminus K_1, I_2\setminus K_2)R(I_3\setminus K_3, I_2\setminus K_2) \\ \cdots R(I_{2\ell-1}\setminus K_{2\ell-1}, I_{2\ell}\setminus K_{2\ell}) \cdot R(I_{2\ell}\setminus K_{2\ell}, I_1\setminus K_1)]. \quad (4.5.4)$$

We now investigate when does a term in the expansion above contribute a non-zero value to the LHS.

First consider the case of $Q = \sum_{s,K} e^i_{s,K}$. Fix $i = 1$, the other cases are similar. $e^1_{s,K}(I,J)$ is a function of the variables g_b for $b \in \mathcal{E}_{ext}(\{s\}, I)$ (whenever $I\cap J = K$). Writing

$e^1_{s,K}(I,J)$ as a polynomial in g_b for $b \in \mathcal{E}_{ext}(\{s\}, I)$ we observe that: $\mathbb{E}_G[e^1_{s,K}] = 0$ and that all coefficients of degree j polynomials are equal for every j and at most 2^d. We decompose the matrix $e^1_{s,K}$ so that for each (I,J) we only pick one of the (corresponding) terms in the polynomial expansion of each entry in g_b described above.

In the expansion of the expected trace above, then, for any such matrix that appears in the decomposition, each term is a (scaled) product of g_b variables for some b. For the expectation of such a term to be non zero, each g_b must occur an even number of times. Consider the case of a matrix in the decomposition of $e^1_{s,K}$ with entries being some (corresponding) monomials of degree 1 for concreteness. The case of other matrices is similar. Fix any term. Let T be the set of all vertices that appear in some I_i for $i \leqslant 2\ell$ and are part of some b for g_b that appears in the term. Then, by a random partitioning argument based on 2, we can first assume that all $\{s_1, s_2, \ldots, s_{2\ell}\}$ doesn't intersect T (and lose a logarithmic factor in the spectral norm upper bound). It is now immediate that every $s \in \{s_1, s_2, \ldots, s_{2\ell}\}$ for a term to have non zero expectation (otherwise, some g_b will not appear twice in the product sequence describing the term). Thus, the number of distinct vertices in any term with a non zero expectation is $\ell + (d-1)2\ell = (d - \frac{1}{2})(2\ell)$. The number of possible terms with the same set of distinct vertices is at most $(2\ell)!$. Finally, each term contributes at most 2^d. Thus, we can upper bound the expected trace of such a matrix by $2^{O(d)} \cdot (2\ell)! \cdot n^{d-\frac{1}{2}}$ poly log (n). We now do the standard step of using the Markov's inequality to obtain an upper estimate on $\mathrm{Tr}((QQ^\dagger)^{2\ell})$ that holds with probability $1 - 1/n$, take $(2\ell)^{th}$ root and finally use $\ell = O(\log(n)$ to obtain the desired bound. Finally, matrix in the decomposition based on polynomial expansion in g_b of the entries of $e^1_{s,K}$ can be similarly upper bounded in spectral norm completing the analysis of this case by an

application of the triangle inequality.

The case of the matrix $\sum_{s\in[n],K:|K|=q} e^i_{s,K} - \tilde{e}^i_{s,K}$ is similar, except that it is now a $(q+1)^{th}$ lift of some matrix. Repeating the reasoning as above, (except each entry (I, J) being described by sets of size $d-2$ instead of $d-1$), we obtain the stated upper bounded in the statement of the lemma.

Finally, we now proceed to the analysis for the third case. We first split the Hadamard product into a sum and analyze each term separately. We sketch the main difference from above for the combinatorial picture for the term $Q = \sum_{s,K} e^1_{s,K} e^2_{s,K}$ here. The other arguments are similar. We again write the $\mathbb{E}[\mathrm{Tr}((QQ^\dagger)^\ell)]$ as a sum over $K_1, \ldots, K_{2\ell}, s_1, s_2, \ldots, s_{2\ell}$ and $I_1, I_2, \ldots, I_{2\ell}$ as above. By expanding out each of $e^1_{s,K}$ and $e^2_{s,K}$ as polynomials in appropriate g_b variables, we observe that the least degree of any term in the expansion is at least 2 (i.e. involves a product of at least 2 g_bs). Decomposing the matrix so that each entry gets the corresponding monomial in the polynomial expansion in terms of g_bs as above, we now consider the matrix with the term involving a scaled product of exactly 2 g_bs. The other matrices in the decomposition can be handled similarly. By a random partitioning argument (2) as before, we can assume that $\{s_1, s_2, \ldots s_{2\ell}\}$ are disjoint from T, the subset of vertices from I_i for $i \leqslant 2\ell$ that appear in b for some g_b and lose a $O(\log(n))$ factor in the estimate on the norm. Again reasoning as before that for a non zero expectation, the term must have each g_b appear an even number of times. Since each g_b must appear at least twice (whenever it appears at least once), the number of distinct g_bs that appear in a term that contributes non zero expectation is at most 2ℓ. On the other hand, we now observe that the $\{s_1, s_2, \ldots, s_{2\ell}\} \cup T$ are all connected via a path using b from g_b that appear in the term and thus, the number of distinct elements in

$\{s_1, s_2, \ldots, s_{2\ell}\} \cup T$ are at most $2\ell + 1$. Thus, a non zero contributing term has a total of at most $2\ell(d-2) + 2\ell = 2\ell(d-1)$. Arguing in the standard way as done above, this now yields a norm estimate of $\tilde{O}(2^{O(d)} \cdot n^{d-1})$ as required.

□

Next, we show find out the spaces $\tilde{e}^i_{s,K}$ contribute to. Towards this, for each i, we define :

$$\tilde{e}^i_{s,q} \stackrel{\text{def}}{=} \sum_{K:|K|=q} \tilde{e}^i_{s,K}.$$

Then, we have:

Lemma 4.5.9. *For any $t, t' > q$, we have:*

1. $$\Pi_t \tilde{e}^2_{s,q} = 0,$$

2. $$\tilde{e}^1_{s,q} \Pi_t = 0,$$

3. $$\Pi_t \tilde{e}^3_{s,q} \Pi_{t'} = 0.$$

Proof. We only do the proof for the first case, the others are similar. The idea is again to use Lemma 4.4.4. Let $v \in \mathbb{R}^{\binom{[n]}{d}}$. We will show that $\tilde{e}^2_{s,q} \cdot v = w$ such that there exists a vector $u \in \mathbb{R}^{\binom{[n]}{q}}$ such that for every I, $w_I = \sum_{I' \subseteq I,\, |I'|=q} u_{I'}$. An application of Lemma 4.4.4 then completes the proof of part (1).

We have:

$$\begin{aligned} w_I &= \sum_{J \in \binom{[n]}{d}} \tilde{e}^2_{s,q}(I,J) v_J \\ &= \sum_{K:|K|=q} \sum_{J \in \binom{[n]}{d}} \tilde{e}^2_{s,K}(I,J) v_J \\ &= \sum_{K:K \subseteq I,\, |K|=q} \sum_{J \in \binom{[n]}{d}} \tilde{e}^2_{s,K}(I,J) v_J. \end{aligned}$$

The proof now follows by observing that $\sum_{J \in \binom{[n]}{d}} \tilde{e}^2_{s,K}(I,J) v_J$ depends only on K. Thus, we set u_K for every $|K| = q$ by $u_K = \sum_{J \in \binom{[n]}{d}} \tilde{e}^2_{s,K}(I,J) v_J$ for any I such that $K \subseteq I$.

□

We can now complete the proof of Lemma 4.5.3 using Lemma 4.5.8and Lemma 4.5.9.

Proof of Lemma 4.5.3. For each q, let A_q be the expression given by Lemma 4.5.7 to approximate the entries of Δ_q. Then, we have for any i, j:

$$\begin{aligned} \Pi_i \Delta \Pi_j &= \sum_{q=1}^{d} \Pi_i \Delta_q \Pi_j \\ &= \sum_{q=1}^{d} \Pi_i \left(\Delta_q - A_q \right) \Pi_j + \sum_{q=1}^{d} \Pi_i A_q \Pi_j \end{aligned}$$

By a simple application of Fact 4.1.1, it is easy to observe that $\|\Delta_q - A_q\| \leqslant 2^{O(d)} \tilde{O}(\omega^{2d-q} \cdot n^{d-1})$. An application of Lemma 4.5.9 and Lemma 4.5.8 yields that the terms

that contribute to $\Pi_i A_q \Pi_j$ have norm at most $2^{O(d)}\tilde{O}(\omega^{2d-\min\{i,j\}} n^{d-\frac{1}{2}}) + 2^{O(d)}\tilde{O}(\omega^{2d-q} n^{d-1})$. This completes the proof.

□

4.6 Analyzing Deviations for the Corrected Degree-4 Operator

The goal of this section is to prove Lemma 4.3.6, that is, to show that $\mathcal{N}' \succ 0$. The proof is organized into 5 main claims that we next present.

We first show that it is enough to prove PSDness of a somewhat simplified matrix $\mathcal{N}$. $\mathcal{N}$ is produced by two simplifications to $\mathcal{N}'$. First, to take care of the zero rows as in Section 4.5, we work with a matrix where we "fill in" the entries carefully. Second, $\mathcal{N}$ and $\mathcal{M}'$ are equal on all entries (I, J) such that $|I \cup J| \leqslant 3$; in other words, the correction affects only the homogeneous degree 4 parts. Specifically, let $\mathcal{R} \in \mathbb{R}^{\binom{[n]}{2} \times \binom{[n]}{2}}$ be defined so that:

$$\mathcal{R}(I,J) = \begin{cases} \gamma(\frac{\omega}{n})^5 C_4 \cdot \sum_{s\in[n]} \prod_{a \in I\cup J} r_s(a) & \text{if } |I \cup J| = 4 \text{ and } \mathcal{E}_{ext}(I,J) \subseteq G \\ 0 & \text{otherwise}\,. \end{cases}$$

(Recall that C_4 is the number of 4-cliques in G.) We then set

$$\mathcal{N} = \mathcal{M}' + \mathcal{R}\,,$$

where $\mathcal{M}'$ is the filled in MPW matrix (Definition 4.2.5).

Our first claim shows that it is enough to prove PSDness of $\mathcal{N}$:

Lemma 4.6.1. *For any γ, c there is $\omega_0 = \Omega(\sqrt{n/\log(n)c\gamma})$ so that for any $\omega \leqslant \omega_0$ with probability $1 - O(n^{-10})$ if $\mathcal{N} \succ \omega^2 n^2/c \cdot I$ then $\mathcal{N}' \succ 0$.*

Next, we decompose the matrix $\mathcal{N}$ appropriately and study the spectrum of each of the pieces. Towards this, we define $\tilde{\mathcal{R}}_0 \in \mathbb{R}^{\binom{[n]}{2} \times \binom{[n]}{2}}$ as follows. For every I, J:

$$\tilde{\mathcal{R}}_0(I,J) = \frac{1}{16} \sum_{s \in [n]} \gamma (\tfrac{\omega}{n})^5 C_4 \cdot \Pi_{a \in (I \cup J) \setminus (I \cap J)} r_s(a)$$

Recall that from Definition 4.5.1, we know that $\mathcal{M}' = E + L + \Delta$. By writing $\mathcal{R} = \tilde{\mathcal{R}}_0 + (\mathcal{R} - \tilde{\mathcal{R}}_0)$, we obtain the decomposition:

$$\mathcal{N} = (E + \tilde{\mathcal{R}}_0) + L + \Delta + (\mathcal{R} - \tilde{\mathcal{R}}_0). \tag{4.6.1}$$

In what follows, we will analyze each piece of the decomposition above separately on a carefully constructed decomposition of $\mathbb{R}^{\binom{[n]}{2}}$. We now proceed and construct this decomposition. Recall $\mathbb{R}^{\binom{[n]}{2}} = V_0 \oplus V_1 \oplus V_2$ where V_0, V_1, V_2 are the eigenspaces of the matrix $E = \mathbb{E}[\mathcal{M}'] = \mathbb{E}[\mathcal{N}]$ from Fact 4.2.9 and Lemma 4.2.10. Let $r_s \in \mathbb{R}^n$ be as described in Definition 4.3.3. With slight abuse of notation, we write $r_s^{\otimes 2}$ for the vector in $\mathbb{R}^{\binom{[n]}{2}}$ such that for every $I \in \binom{[n]}{2}$,

$$r_s^{\otimes 2}(I) = \Pi_{i \in I} r_s(i).$$

We now define a new decomposition by splitting V_2 further and write $\mathbb{R}^{\binom{[n]}{2}} = W_0 \oplus W_1 \oplus W_{1.5} \oplus W_2$, where:

$$\begin{aligned}
&W_0 = V_0\,, \\
&W_1 = V_1\,, \\
&W_{1.5} \text{ s.t. } W_{1.5} \perp (W_0 \oplus W_1) \text{ and } W_0 \oplus W_1 \oplus W_2 = V_0 \oplus V_1 \oplus \mathsf{Span}\{r_s^{\oplus 2}\} \\
&W_2 = (W_1 \oplus W_2 \oplus W_3)^{\perp}\,.
\end{aligned} \tag{4.6.2}$$

Let Π_{W_a} be the projector to W_a for every $a \in \{0, 1, 1.5, 2\}$.

We are now ready to analyze the spectrum of each piece from (4.6.1). First, we analyze the spectrum of $(E + \tilde{\mathcal{R}}_0)$:

Lemma 4.6.2. *For every γ there is $\omega_0 = \Omega(\sqrt{\gamma n}/\log(n)^2)$ so that for $\omega \leqslant \omega_0$, with probability $1 - O(n^{-10})$,*

$$(E + \tilde{\mathcal{R}}_0) \succeq \Omega(\omega^4 n^2) \cdot \Pi_{W_0} + \Omega(\omega^3 n^2) \cdot \Pi_{W_1} + \Omega(\gamma \omega^5 n) \cdot \Pi_{W_{1.5}} + \Omega(\omega^2 n^2) \cdot \Pi_{W_2}.$$

Next, we analyze the spectrum of L. Here at last we see the main technical improvement in these corrected moments—the cross term between V_1 and V_2 has become a cross term between W_1 and W_2 and has much-reduced norm.

Lemma 4.6.3. *For any $\omega = \tilde{O}(\sqrt{n})$ there is $p = \operatorname{polylog} n$ so that, with probability $1 - O(n^{-10})$ the following bounds hold.*

1. *Diagonal Terms:*

$$\|\Pi_{W_a} L \Pi_{W_a}\| \leqslant O(p\omega^4 n^{3/2}) \quad \text{for } a \in \{0, 1, 1.5\}$$
$$\|\Pi_{W_2} L \Pi_{W_2}\| \leqslant O(p\omega^4 n)\,.$$

2. *Off-Diagonal Terms:*

$$\|\Pi_{W_a} L \Pi_{W_b}\| \leqslant O(p\omega^4 n^{3/2}) \quad \text{for } a, b \in \{0, 1, 1.5, 2\}$$
$$\|\Pi_{W_a} L \Pi_{W_2}\| \leqslant O(p\omega^3 n^{3/2}) \quad \text{for } a \in \{1, 1.5\}\,.$$

Next, we bound the spectral norm of Δ. This is a direct corollary of the more general bound in Lemma 4.5.3, but to have a self-contained proof of the degree-4 case we also give a proof later in this section.

Lemma 4.6.4. *Let $\mathcal{M}'$ be the $\binom{n}{2} \times \binom{n}{2}$ filled-in matrix for the degree-4 MPW moments with clique size ω (Definition 4.2.5). Let Δ be as in Definition 4.5.1. With probability $1 - O(n^{-10})$, $\|\Delta\| \leqslant O(\omega^3 n^{3/2} \log(n)^2)$.*

Finally, we bound the spectral norm of the last piece $(\mathcal{R} - \tilde{\mathcal{R}}_0)$:

Lemma 4.6.5. *Let $G \sim G(n, 1/2)$. With probability $1 - O(n^{-10})$, $\|\mathcal{R} - \mathcal{R}_0\| \leqslant O(\gamma\omega^5 n^{1/2} \log(n)^2)$.*

The proofs of these lemmas follow, but first we complete the proof of Lemma 4.3.6 and hence of Theorem 4.3.4.

Proof of Lemma 4.3.6. By Lemma 4.6.1, it will be enough to exhibit $c, \gamma = \operatorname{polylog} n$ and $\omega_1 = \Omega(\sqrt{n/\log(n)c\gamma})$ so that $\mathcal{N} > (\omega^2 n^2/c) \cdot I$ with probability $1 - O(n^{-10})$ when $\omega \leqslant \omega_1$. Then our final bound will be given by the minimum of ω_1 and ω_0 of Lemma 4.6.1. (Recall that γ is a parameter inside $\mathcal{N}$.) In the following, all that we claim happens with probability at least $1 - n^{-9}$ by a union bound.

So let $\omega_0 \in \mathbb{R}$; we will choose it later. We will find c, γ and conditions on ω_0 so that the conditions of Lemma 4.1.2 hold for $\mathcal{N} - (\omega^2 n^2/c) \cdot I$.

First of all, by Lemma 4.6.5, for every γ there is $c = O(\min\{n/\omega^3\gamma, 1\})$ so that

$$E - (\omega^2 n^2/c) \cdot I \geq \Omega(\omega^4 n^2) \cdot \Pi_{W_0} + \Omega(\omega^3 n^2) \cdot \Pi_{W_1} + \Omega(\gamma\omega^5 n) \cdot \Pi_{W_{1.5}} + \Omega(\omega^2 n^2) \cdot \Pi_{W_2}\,.$$

We assume $c = c(\gamma)$ is chosen in this way.

For any γ, by Lemma 4.6.5 if we choose $\omega \leqslant \sqrt{n}/\gamma \log(n)^2$ then $\|\mathcal{R} - \mathcal{R}_0\| \leqslant o(\omega^2 n^2)$. Adding $\mathcal{R} - \mathcal{R}_0$ to the previous equation,

$$(E + \mathcal{R}_0) + (\mathcal{R} - \mathcal{R}_0) - (\omega^2 n^2/c) \cdot I \geq \Omega(\omega^4 n^2) \cdot \Pi_{W_0} + \Omega(\omega^3 n^2) \cdot \Pi_{W_1} + \Omega(\gamma\omega^5 n) \cdot \Pi_{W_{1.5}} + \Omega(\omega^2 n^2) \cdot \Pi_{W_2}\,.$$

By the same reasoning using Lemma 4.6.4 to add Δ to the previous equation, we have for the same choice of ω:

$$\mathcal{N} - L - (\omega^2 n^2/c)\cdot I \succeq \Omega(\omega^4 n^2)\cdot\Pi_{W_0} + \Omega(\omega^3 n^2)\cdot\Pi_{W_1} + \Omega(\gamma\omega^5 n)\cdot\Pi_{W_{1.5}} + \Omega(\omega^2 n^2)\cdot\Pi_{W_2}\,. \quad (4.6.3)$$

So it just remains to add L to the left-hand side.

We decompose L as:

$$L = (\Pi_{W_0} + \Pi_{W_1} + \Pi_{W_{1.5}})L(\Pi_{W_0} + \Pi_{W_1} + \Pi_{W_{1.5}}) + \Pi_{W_2}L + L\Pi_{W_2}\,.$$

Let p be as in Lemma 4.6.3, which implies that

$$(\Pi_{W_0} + \Pi_{W_1} + \Pi_{W_{1.5}})L(\Pi_{W_0} + \Pi_{W_1} + \Pi_{W_{1.5}}) \preceq O(p\omega^4 n^{3/2})(\Pi_{W_0} + \Pi_{W_1} + \Pi_{W_{1.5}})\,. \quad (4.6.4)$$

Choosing $\gamma = O(p^2\log(n))$ we get that this is $o(\sqrt{\omega^4 n^2 \cdot \gamma\omega^5 n})$. So using Lemma 4.1.3 to add (4.6.3) and (4.6.4), we obtain

$$\mathcal{N} - \Pi_{W_2}L - L\Pi_{W_2} - (\omega^2 n^2/c)\cdot I \succeq \Omega(\omega^4 n^2)\cdot\Pi_{W_0} + \Omega(\omega^3 n^2)\cdot\Pi_{W_1} + \Omega(\gamma\omega^5 n)\cdot\Pi_{W_{1.5}} + \Omega(\omega^2 n^2)\cdot\Pi_{W_2}\,. \quad (4.6.5)$$

We break $\Pi_{W_2}L$ apart as

$$\Pi_{W_2}L = \Pi_{W_2}L\Pi_{W_0} + \Pi_{W_2}L\Pi_{W_2} + \Pi_{W_2}L(\Pi_{W_1} + \Pi_{W_{1.5}})\,.$$

By Lemma 4.6.3,

$$\begin{aligned}
\|\Pi_{W_2}L\Pi_{W_0}\| &\leqslant O(p\omega^4 n^{3/2}) = o(\sqrt{\omega^4 n^2 \cdot \omega^2 n^2}) && \text{for } \omega \leqslant \sqrt{n}/p\log(n)\\
\|\Pi_{W_2}L\Pi_{W_2}\| &\leqslant O(p\omega^4 n) = o(\omega^2 n^2) && \text{for } \omega \leqslant \sqrt{n}/p\log(n)\\
\|\Pi_{W_2}L(\Pi_{W_1} + \Pi_{W_{1.5}})\| &\leqslant O(p\omega^3 n^{3/2}) = o(\sqrt{\omega^2 n^2 \cdot \gamma\omega^5 n}) && \text{for } \gamma \geqslant p^2\log p\,.
\end{aligned}$$

Together with Lemma 4.1.3 and (4.6.5), this implies the lemma, for $\gamma \geqslant p^2\log p$, $c = c(\gamma)$ as above, and $\omega_0 \leqslant \min\{\sqrt{n}/p\log(n)^2, \sqrt{n}/\gamma\log(n)^2\}$. □

4.6.1 Proof of Diagonal and Off-Diagonal Norm Bounds (Lemma 4.6.3)

Here we prove Lemma 4.6.3.

Proof of Lemma 4.6.3. We start with the easy parts. Note that $\Pi_{W_2} L \Pi_{W_2} = \Pi_{W_2}\Pi_{V_2} L \Pi_{V_2}\Pi_{W_2}$, so the bound

$$\|\Pi_{W_2} L \Pi_{W_2}\| \leqslant \tilde{O}(\omega^4 n)$$

is immediate from Lemma 4.5.2. The same theorem also implies that $\|L\| \leqslant \tilde{O}(\omega^4 n^{3/2})$, and since projectors are contractive this finishes the bounds on the diagonal terms and the first part of the off-diagonal bound (for cross terms among $W_0, W_1, W_{1.5}$).

We just have to prove that $\|\Pi_{W_1} L \Pi_{W_2}\| \leqslant \tilde{O}(\omega^4 n)$. We will use the patterned matrix machinery to show this. We recall the decomposition of L from Section 4.5 as $L = L_0 + L_1 + L_2$. The main point is to show that $\|\Pi_{W_1} L_0 \Pi_{W_2}\| \leqslant \tilde{O}(\omega^4 n)$, so we postpone to the end of the proof the case of L_1 and L_2.

Recall again that L_0 can be further decomposed as $L_0 = O(1) \cdot \sum_{B \in \mathcal{B}} L_0^B$ (again, see Section 4.5 for definitions). Finally, for each L_0^B there is a $\tilde{L}_0^B$ so that $\|L_0^B - \tilde{L}_0^B\| \leqslant \tilde{O}(\omega^4 n)$ with probability $1 - O(n^{-10})$ (see Claim 4.5.4). Since $|\mathcal{B}| = O(1)$, it is actually enough for us to show that

$$\left\| \Pi_{W_1} \left(\sum_{B \in \mathcal{B}} \tilde{L}_0^B \right) \Pi_{W_2} \right\| \leqslant \tilde{O}(\omega^4 n)\,.$$

Let $\mathcal{B}_1$ be the bipartite graphs on $[2] \times [2]$ for which at least one right-hand vertex has degree 0. Then by Lemma 4.4.9, $\sum_{B \in \mathcal{B}} \tilde{L}_0^B \Pi_{W_2} = 0$, since $W_2 \subseteq V_2$.

We are left with $\mathcal{B}_2$, the family of bipartite graphs where every right-hand-side vertex has nonzero degree. Using Lemma 4.4.11 on the matrices L_0^B, we get that if B has

no vertices of degree 0 then $\|L_0^B\| \leqslant \tilde{O}(\omega^4 n)$ with probability $1 - O(n^{-10})$. Since as above $\|L_0^B - \tilde{L}_0^B\| \leqslant \tilde{O}(\omega^4 n)$ with similar probability, we get $\|\tilde{L}_0^B\| \leqslant \tilde{O}(\omega^4 n)$ for these B.

The only remaining graphs in $\mathcal{B}$ are the two graphs B_1, B_2 with exactly one vertex on the left-hand side of degree 0. It is not hard to check that rows $\{a, b\}$ of the matrices $\tilde{L}_0^{B_1}$ and $\tilde{L}_0^{B_2}$ are matrices are $r_a^{\otimes 2}$ and $r_b^{\otimes 2}$, respestively. Since $W_2 \perp r_s^{\otimes 2}$, we get $\tilde{L}_0^{B_1}\Pi_{W_2} = \tilde{L}_0^{B_2}\Pi_{W_2} = 0$.

It remains to handle L_1 and L_2. By Claim 4.5.5 together with Claim 4.5.4, each satisfies $\|L_1\|, \|L_2\| \leqslant \tilde{O}(\omega^3 n^{3/2})$. Since $\omega \leqslant \sqrt{n}$, the lemma now follows. □

4.6.2 Lower-Degree Cleanup (Lemma 4.6.1)

In this section we prove Lemma 4.6.1. We start by bounding the difference between our pseudoexpectation $\tilde{\mathbb{E}}$ and the MPW operator on polynomials of degree less than 4. This lemma is a consequence of Lemma 4.3.7 and the Gershgorin circle theorem (Fact 4.1.1).

Lemma 4.6.6. *Let $G \sim G(n, 1/2)$. Let ω be a real parameter. Let $\tilde{\mathbb{E}}$ be as given in Definition 4.3.3. Let $\tilde{\mathbb{E}}_0$ be the MPW operator for clique size ω. Suppose $\omega \leqslant \sqrt{n}$ and $\gamma = O(\omega)$. Then with probability at least $1 - O(n^{-20})$,*

- *Every $i, j, k \in [n]$ with i, j, k all distinct satisfies*

$$|\tilde{\mathbb{E}}\, x_i x_j x_j - \tilde{\mathop{\mathbb{E}}_0}\, x_i x_j x_k| \leqslant O(\gamma \log(n)\omega^4/n^4)\,.$$

- *Every $i, j \in [n]$ with $i \neq j$ satisfies $|\tilde{\mathbb{E}}\, x_i x_j - \tilde{\mathbb{E}}_0\, x_i x_j| \leqslant o(\omega^2/n^2)$.*

Proof of Lemma 4.6.6. Recall that we obtained $\tilde{\mathbb{E}}$ by starting with the clique-size ω MPW operator on multilinear homogeneous degree-4 polynomials, adding a correction operator

on those same polynomials, and then infering values of $\tilde{\mathbb{E}}$ on lower-degree polynomials via the constraint $\sum_i x_i = \omega'$. There are two primary sources of the difference between our operator $\tilde{\mathbb{E}}$ and the MPW operator on polynomials of lower degree. The dominant one is the propogation to lower degree polynomials of the correction operator $\mathcal{L}$ (recall Definition 4.3.1). The second is that the degree-4 values coming from the MPW part of our operator $\tilde{\mathbb{E}}$ are propogated downwards using the constraint $\sum_i x_i = \omega'$ rather than $\sum_i x_i = \omega$ as is done to define the rest of the MPW operator.

We start with the degree 3 bound. Denote by $\tilde{\mathbb{E}}_0$ the MPW operator for clique size ω. Consider $i, j, k \in [n]$ all distinct, and recall that

$$\begin{aligned}
\tilde{\mathbb{E}}\, x_i x_j x_k &= \frac{1}{\omega' - 3} \sum_{\ell \neq i,j,k} \tilde{\mathbb{E}}\, x_i x_j x_k x_\ell \\
&= \frac{1}{\omega' - 3} \sum_{\ell \neq i,j,k} \left(\tilde{\mathop{\mathbb{E}}_0}\, x_i x_j x_k x_\ell + \mathcal{L} x_i x_j x_k x_\ell \right) \\
&= \left(\frac{1}{\omega' - 3} - \frac{1}{\omega - 3} \right) \left(\sum_{\ell \neq i,j,k} \tilde{\mathop{\mathbb{E}}_0}\, x_i x_j x_k x_\ell \right) + \tilde{\mathop{\mathbb{E}}_0}\, x_i x_j x_k + \frac{1}{\omega' - 3} \sum_{\ell \neq i,j,k} \mathcal{L} x_i x_j x_k x_\ell \,.
\end{aligned}$$

Thus,

$$\tilde{\mathbb{E}}\, x_i x_j x_k - \tilde{\mathop{\mathbb{E}}_0}\, x_i x_j x_k = \left(\frac{1}{\omega' - 3} - \frac{1}{\omega - 3} \right) \left(\sum_{\ell \neq i,j,k} \tilde{\mathop{\mathbb{E}}_0}\, x_i x_j x_k x_\ell \right) + \frac{1}{\omega' - 3} \sum_{\ell \neq i,j,k} \mathcal{L} x_i x_j x_k x_\ell \,.$$

By Lemma 4.3.7, with probability $1 - O(n^{-25})$ every i, j, k satisfies $|\sum_{\ell \neq i,j,k} \mathcal{L} x_i x_j x_k x_\ell| \leqslant O(\gamma \omega^5 \log(n)/n^4)$. At the same time, we know by Lemma 4.3.5 together with Lemma 4.3.7 that $|\omega' - \omega| \leqslant O(\gamma \omega^2 \log(n)^2/n^{5/2})$. This implies that $|1/(\omega' - 3) - 1/(\omega - 3)| \leqslant O(\gamma \log(n)^2/n^{5/2})$ (all with probability at least $1 - O(n^{-25})$). In conjunction with the preceeding, it implies also that every i, j, k satisfies $(1/(\omega' - 3))|\sum_{\ell \neq i,j,k} \mathcal{L} x_i x_j x_k x_\ell| \leqslant$

$O(\gamma \log(n)\omega^4/n^4)$. Together with the trivial bound $|\sum_{\ell \neq i,j,k} \tilde{\mathbb{E}}_0 x_i x_j x_k x_\ell| \leqslant O(\omega^4/n^3)$ with probability $1 - n^{-\omega(1)}$ (following from [85, Theorem 10.3]), all this implies that with probability $1 - O(n^{-25})$,

$$|\tilde{\mathbb{E}} x_i x_j x_k - \tilde{\mathop{\mathbb{E}}_0} x_i x_j x_k| \leqslant O(\gamma \log(n)^2 \omega^4/n^{11/2}) + \tilde{O}(\gamma \log(n)\omega^4/n^4) = \tilde{O}(\gamma \log(n)\omega^4/n^4)\,.$$

We turn to the degree-two bound. Fix $i \neq j \in [n]$. We expand $\tilde{\mathbb{E}} x_i x_j - \tilde{\mathbb{E}}_0 x_i x_j$.

$$\begin{aligned}
&\tilde{\mathbb{E}} x_i x_j - \tilde{\mathop{\mathbb{E}}_0} x_i x_j \\
&= \frac{1}{(\omega'-2)(\omega'-3)} \sum_{k \neq i,j} \tilde{\mathbb{E}} x_i x_j x_k - \frac{1}{(\omega-2)(\omega-3)} \sum_{k \neq i,j} \tilde{\mathop{\mathbb{E}}_0} x_i x_j x_k \\
&= \frac{1}{(\omega'-2)(\omega'-3)} \sum_{k \neq i,j} (\tilde{\mathbb{E}} x_i x_j x_k - \tilde{\mathop{\mathbb{E}}_0} x_i x_j x_k) - \\
&\left(\frac{1}{(\omega-2)(\omega-3)} - \frac{1}{(\omega'-2)(\omega'-3)}\right) \sum_{k \neq i,j} \tilde{\mathop{\mathbb{E}}_0} x_i x_j x_k\,.
\end{aligned}$$

With probability $1 - O(n^{-20})$ when $G \sim G(n, 1/2)$ by Lemma 4.3.7, we get that $1/(\omega'-2)(\omega'-3) = O(1/\omega^2)$. By the same,

$$\left|\frac{1}{(\omega-2)(\omega-3)} - \frac{1}{(\omega'-2)(\omega'-3)}\right| \leqslant \tilde{O}(\gamma/\omega n^{5/2})\,.$$

Together with the bound from earlier in this proof on $|\tilde{\mathbb{E}} x_i x_j x_k - \tilde{\mathbb{E}}_0 x_i x_j x_k|$ and the trivial bound $|\sum_{k \neq i,j} \tilde{\mathbb{E}}_0 x_i x_j x_k| \leqslant \tilde{O}(\omega^3/n^2)$, we obtain

$$|\tilde{\mathbb{E}} x_i x_j - \tilde{\mathop{\mathbb{E}}_0} x_i x_j| \leqslant \tilde{O}(\gamma\omega^2/n^3) + \tilde{O}(\gamma\omega^2/n^{9/2})$$

which is $o(\omega^2/n^2)$ for $\omega = o(\sqrt{n})$ and $\gamma = o(\omega)$. □

Proof of Lemma 4.6.1. We first observe that an eigenvalue lower bound on $\mathcal{N}$ implies the same on the (principal) submatrix indexed only by cliques (in this case, edges) in G. This submatrix is equal to $C_4\mathcal{N}' + \mathrm{Err}$, where

$$\mathrm{Err}(I,J) = \begin{cases} C_4 \tilde{\mathbb{E}}\, x^I x^J - C_4 \tilde{\mathbb{E}}_0\, x^I x^J & \text{if } |I \cup J| < 4 \\ 0 & \text{otherwise} \end{cases}$$

We break Err into two parts so that $\mathrm{Err}_2 + \mathrm{Err}_3 = \mathrm{Err}$:

$$\mathrm{Err}_2(I,J) = \begin{cases} \mathrm{Err}(I,J) & \text{if } |I \cup J| = 2 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \mathrm{Err}_3(I,J) = \begin{cases} \mathrm{Err}(I,J) & \text{if } |I \cup J| = 3 \\ 0 & \text{otherwise} \end{cases}$$

Note that Err_3 consists of the off-diagonal nonzero entries of Err, while Err_2 contains the diagonal entries of Err. We start by showing a bound on $\|\mathrm{Err}_3\|$.

$$\begin{aligned} \|\mathrm{Err}_3\| &\leqslant \max_{I \in \binom{n}{2}} \sum_{J \neq I} |\mathrm{Err}_3(I,J)| \\ &= \max_{\{i_1,i_2\} \in \binom{n}{2}} \sum_{k \notin \{i_1,i_2\}} |\mathrm{Err}_3(I,\{k,i_1\})| + |\mathrm{Err}_3(I,\{k,i_2\})| \quad \text{(definition of } \mathrm{Err}_3\text{)} \\ &\leqslant O(n) \cdot \max_{i,j,k \text{ all not equal}} C_4 |\tilde{\mathbb{E}}\, x_i x_j x_k - \tilde{\mathbb{E}}_0\, x_i x_j x_k| \\ &\leqslant O(n) \cdot C_4 \cdot O(\gamma \log(n) \omega^4 / n^4) \quad \text{w.p. } 1 - O(n^{-20}) \text{ by Lemma 4.6.6} \\ &\leqslant O(\gamma \log(n) \omega^4 n) \quad \text{w.p. } 1 - O(n^{-20}) \text{ by } C_4 \approx n^4\text{, see [85, Theorem 10.3]} \end{aligned}$$

Next we bound Err_2. Since it is diagonal, it is enough to give an entrywise bound.

$$\begin{aligned} \|\mathrm{Err}_2\| &\leqslant \max_{I \in \binom{n}{2}} \mathrm{Err}_2(I,I) \\ &= \max_{i \neq j} C_4(\tilde{\mathbb{E}}\, x_i x_j - \tilde{\mathbb{E}}_0\, x_i x_j) \quad \text{by definition of } \mathrm{Err}_2 \\ &\leqslant C_4 \cdot o(\omega^2/n^2) \quad \text{w.p. } 1 - O(n^{-20}) \text{ by Lemma 4.6.6} \\ &\leqslant o(\omega^2 n^2) \quad \text{w.p. } 1 - O(n^{-20}) \text{ by } C_4 \approx n^4\text{, see [85, Theorem 10.3]}\,. \end{aligned}$$

Fix $\gamma, c \in \mathbb{R}$. Suppose $\mathcal{N} \geq (\omega^2 n^2/c) \cdot I$. Then for $\mathcal{N} + \mathrm{Err}$ to be PSD it is enough to have $\|\mathrm{Err}\| \leqslant \omega^2 n^2/c$. There is by the above bounds a universal constant C so that it is enough to have $Cc\gamma \log(n)\omega^4 n \leqslant \omega^2 n^2$, or rearranging, $\omega \leqslant \sqrt{n/Cc\log(n)\gamma}$. □

4.6.3 Eigenvalue Lower Bound for the Correction

The following is the main claim for this section.

Lemma 4.6.7. *Let $G \sim G(n, 1/2)$. Let*

$$r_s(j) = \begin{cases} 1 & \text{if } s \sim j \\ -1 & \text{if } s \nsim j \\ 0 & \text{if } s = j \end{cases}.$$

Let $r_s^{\otimes 2} \in \mathbb{R}^{\binom{[n]}{2}}$ be the vector with entries $r_s^{\otimes 2}(\{i,j\}) = r_s(i) r_s(j)$. Let $V = \mathsf{Span}\{r_s^{\otimes 2}\}_{s \in [n]}$. Let Π_V be the projector to V. With probability at least $1 - O(n^{-10})$ over the sample of G,

$$\sum_s (r_s^{\otimes 2})(r_s^{\otimes 2})^\top \geq \Omega(n^2)\Pi_V .$$

We will need the following graph theoretic machinery for the moment method bound.

Definition 4.6.8. Let G be a graph on $[n]$. A diamond ribbon R of length 2ℓ is a graph on vertices $s_1, \ldots, s_\ell, t_1, \ldots, t_\ell, u_1, \ldots, u_{2\ell}, v_1, \ldots, v_{2\ell}$. It has edges

$$(s_i, u_{2i}), (s_i, v_{2i}), (u_{2i}, t_i), (v_{2i}, t_i), (t_i, u_{2i+1}), (t_i, v_{2i+1}), (u_{2i+1}, s_{i+1}), (v_{2i+1}, s_{i+1})$$

where addition is modulo 2ℓ.

A labeled diamond ribbon (R, F) is a a diamond ribbon of length 2ℓ together with a labeling $F : R \to G$ of vertices in R with vertices from G. We insist that for $(x, y) \in R$ an edge that $F(x) \neq F(y)$.

The labeled diamond ribbon (R, F) is contributing if no element of the multiset $\{(F(x), F(y)) \text{ such that } (x, y) \in R\}$ occurs with odd multiplicity. It is *disjoint* if the sets

$$\{F(s_i)\}, \{F(t_i)\}, \{F(u_i)\}_{i \text{ odd}}, \{F(v_i)\}_{i \text{ even}}, \{F(v_i)\}_{i \text{ odd}}, \{F(v_i)\}_{i \text{ even}}$$

are disjoint.

Lemma 4.6.9. *Let (R, F) be a contributing disjoint labeled diamond ribbon of length 2ℓ. Then it contains at most $3\ell + O(1)$ distinct labels.*

Proof. By our disjointness assumption, every element of the multiset $\{F(s_i), F(t_i)\}$ must occur with multiplicity at least two and similarly for $\{F(u_i)\}$ and $\{F(v_i)\}$. □

Proof of Lemma 4.6.7. Note that the matrix $R = \sum_s (r_s^{\otimes 2})(r_s^{\otimes 2})^\top$ has row and column spaces both V. Note also that it factors as $SS^\top$, where S is the $\binom{[n]}{2} \times n$ matrix whose columns are the vectors $r_s^{\otimes 2}$. Thus, it will be enough to show that

$$S^\top S \geq \Omega(n^2) I \quad \text{w.p. } 1 - O(n^{-10})$$

where here I is the $n \times n$ identity matrix.

For this, consider the matrix $S^\top S$ indexed by vertices $s, t \in [n]$. It has entries

$$S^\top S(s, t) = \langle r_s^{\otimes 2}, r_t^{\otimes 2} \rangle$$

and in particular,

$$S^\top S(s,s) = \langle r_s^{\otimes 2}, r_s^{\otimes 2} \rangle = \binom{[n]}{2}$$

So, zeroing this matrix on the diagonal, it is enough to prove that

$$\left\| S^\top S - \binom{[n]}{2} I \right\| \leqslant o(n^2) \quad \text{w.p. } 1 - O(n^{-10}) .$$

Let $H := S^\top S - \binom{[n]}{2} I$. Let $H_{i,j}$ for $i \neq j$ be given by

$$H_{i,j}(s,t) = \begin{cases} r_s(i) r_s(j) r_t(i) r_t(j) & \text{if } s \neq t \\ 0 & \text{otherwise} \end{cases} .$$

Then $H = \sum_{i \neq j} H_{i,j}$. Note that $H_{i,j}(s,t) = 0$ if $i \in \{s,t\}$ or $j \in \{s,t\}$. Thus the obvious generalization of Lemma 4.4.14 to the two-parameter family $H_{i,j}$ applies. This gives us a family of matrices $H^1, \dots, H^r$ for some $r = O(\log n)$ and a corresponding family of partitions $(S_1^1, S_2^2, S_3^1, S_4^1, S_5^1, S_6^1), \dots, (S_1^r, \dots, S_6^r)$ of $[n]$.

These will be such that $\sum_{i=1}^r H^i = H$, and

$$H^i(s,t) = \sum_{(j,k) \in S_i{}'} r_s(j) r_s(k) r_t(j) r_t(k)$$

where $S^{i'} \subseteq S_5^i \times S_6^i$ is the subset giving indices so that the corresponding summand has not occurred in any $i'' < i$. We will bound $\mathbb{E} \operatorname{Tr}(H^i)^{2\ell}$ for some ℓ to be chosen later. Every term in the expansion of this quantity corresponds to a disjoint labeled diamond ribbon of length 2ℓ, and the number of nonzero terms is at most the number of contributing disjoint labeled diamond ribbons. So $\mathbb{E} \operatorname{Tr}(H^i)^{2\ell} \leqslant n^{3\ell + O(1)}$. The rest follows by standard manipulations. □

4.6.4 Proofs of Remaining Lemmas

Proof of Lemma 4.6.2. By Lemma 4.2.10,

$$E \succeq \Omega(\omega^4 n^2) \cdot \Pi_{W_0} + \Omega(\omega^3 n^2) \cdot \Pi_{W_1} + \Omega(\omega^2 n^2) \cdot \Pi_{W_2} .$$

Let $W = W_0 \oplus W_1$. By Lemma 4.6.7 and [85, Theorem 10.3] (saying that $C_4 \approx n^4$), with probability $1 - O(n^{-10})$ we get that $\tilde{\mathcal{R}}_0 \succeq \Omega(\gamma\omega^5 n)\Pi_{\mathsf{Span}\{r_s^{\otimes 2}\}}$. We make the observation that $\tilde{\mathcal{R}}_0 = (\Pi_W + \Pi_{W_{1.5}})\tilde{\mathcal{R}}_0(\Pi_W + \Pi_{W_{1.5}})$ and that $\Pi_{W_{1.5}}\Pi_{\mathsf{Span}\{r_s^{\otimes 2}\}}\Pi_{W_{1.5}} = \Pi_{W_{1.5}}$. So we just need to handle the term $\Pi_W \tilde{\mathcal{R}}_0 \Pi_{W_{1.5}} + \Pi_{W_{1.5}} \tilde{\mathcal{R}}_0 \Pi_W$.

Together, Lemma 4.4.12 and Lemma 4.3.7 imply that $\|\tilde{\mathcal{R}}_0\| \leqslant O(\gamma\omega^5 n \log(n)^2)$ with probability $1 - O(n^{-10})$. Thus to ensure that $\|\tilde{\mathcal{R}}_0\| \leqslant o(\sqrt{\omega^3 n^2} \cdot \gamma\omega^5 n)$ it is enough to choose $\omega \leqslant \sqrt{\gamma n}/\log(n)^2$. This is enough to apply Lemma 4.1.3 and conclude the proof. □

Proof of Lemma 4.6.4. Note that for I, J disjoint we have $\Delta(I, J) = 0$. We bound the maximal sum across any row of Δ. With probability $1 - O(n^{-10})$ every off-diagonal entry off Δ is at most $O(\omega^3\sqrt{n}\log(n)^2)$ in absolute value [85, Theorem 10.1]. For each $I \in \binom{[n]}{2}$, we then get $\sum_{J \neq I} |\Delta(I, J)| \leqslant O(\omega^3 n^{3/2}\log(n)^2)$. At the same time, The diagonal entries are each at most $O(\omega^2 n^{3/2}\log(n)^2)$ with similar probability, again by [85, Theorem 10.1]. □

Proof of Lemma 4.6.5. $\mathcal{R}$ and $\tilde{\mathcal{R}}_0$ differ in two respects. We first bound the spectral norm of the part of $\tilde{\mathcal{R}}_0$ on non-disjoint entries. Let

$$\tilde{\mathcal{R}}_0^{(3)}(I, J) = \begin{cases} \tilde{\mathcal{R}}_0(I, J) & \text{if } |I \cap J| = 3 \\ 0 & \text{otherwise} \end{cases} .$$

It follows from Lemma 4.3.7 and a row-sum bound that $\|\tilde{\mathcal{R}}_0^{(3)}\| \leqslant O(\gamma\omega^5\sqrt{n}\log(n)^2)$ with probability $1 - O(n^{-10})$. A similar analysis holds for the analogous matrix $\tilde{\mathcal{R}}_0^{(2)}$.

Let $\mathcal{R}_0$ be given by

$$\mathcal{R}_0(I,J) = \begin{cases} \frac{1}{16}\sum_{s\in[n]} \gamma(\frac{\omega}{n})^5 C_4 \cdot \Pi_{a\in(I\cup J)\setminus(I\cap J)} r_s(a) & \text{if } |I\cap J| = 4 \\ 0 & \text{otherwise} \end{cases}.$$

Now it is enough to bound $\|\mathcal{R}_0 - \mathcal{R}\|$. This is the deviation introduced by zeroing non-clique entries. Note that each entry I, J of $\mathcal{R}$ can be decomposed as a function of the underlying graph edge variables:

$$\mathcal{R}(I,J) = \mathcal{R}_0 + \gamma C_4 (\tfrac{\omega}{n})^5 \frac{1}{16} \sum_{\text{nonempty } S\subseteq \mathcal{E}_{ext}(I,J)} \sum_s \prod_{(u,v)\in S} g_{u,v} \prod_{u\in I\cup J} r_s(u)\,,$$

where we recall that for an edge (u,v), the variable $g_{(u,v)}$ is the ± 1 indicator for that edge. Each of the entries in the sum over nonempty S above corresponds to a matrix of the form bounded in Lemma 4.4.12, where we conclude that each has spectral norm at most $O(n^{3/2}\log(n)^2)$ with probability $1 - O(n^{-10})$. We conclude (also using $C_4 \approx n^4$, see [85, Theorem 10.3]) that $\|\mathcal{R}_0 - \mathcal{R}\| \leqslant O(\gamma\omega^5\sqrt{n}\log(n)^2)$ as desired. □

4.7 Concentration of $\deg_G(I)$

In this section, we prove the following large-deviation bounds on the number of x-cliques a random $G(n,1/2)$ graph contains and on $\deg_G(I)$. Similar results (which are likely sufficient for our needs) appear in the literature; see [94, 106, 69] for instance. We provide these proofs for completeness. A coarser concentration result for $\deg_G(I)$ appears in [85].

Definition 4.7.1. For a graph G, define $N_x(G)$ to be the number of x-cliques in G.

Unless otherwise specified, in this section $G \sim G(n,1/2)$.

This first theorem gives the large deviation bound for the number of cliques of size x in G.

Theorem 4.7.2. *For all $\varepsilon \in (0,1)$, for all x, if $n > x^2(2e - e\ln\varepsilon)(2e + 2 - e\ln\varepsilon)$ then*

$$\mathbb{P}\left[\left|N_x(G) - 2^{-\binom{x}{2}}\binom{n}{x}\right| > e(2-\ln\varepsilon)\frac{x^2}{x!}n^{x-1}\right] < \varepsilon.$$

(Note that $2^{-\binom{x}{2}}\binom{n}{x} = \mathbb{E}\,N_x(G)$.)

We also want a large deviations inequality for the number of cliques of size $2d$ that a clique of size $d' < 2d$ participates in in G. Moreover, to carry out the eigenspace splitting arguments needed for Lemma 4.5.3, we want to know the dependence of this deviation on the number of vertices adjacent to every vertex in the d'-clique. The following theorem serves both these purposes.

Definition 4.7.3. Given any $I \subseteq [n]$, let A_I be the set of all vertices not in I which are adjacent to all vertices in I.

Theorem 4.7.4. *There is a universal constant C so that for any $I \subseteq [n]$ of size at most $2d$, if I is a clique in G then for any $\varepsilon \in (0,1)$, if $n \geqslant 100d^2 2^{2d}(3-\ln\varepsilon)^2$ then*

$$\mathbb{P}\left[\left|\deg_G(I) - 2^{\binom{|I|}{2}-\binom{2d}{2}}\binom{n-|I|}{2d-|I|}\right| > C(3-\ln\varepsilon)n^{d-|I|-\frac{1}{2}}\right] < \varepsilon$$

More precisely, if $|I| < d$ then

$$\mathbb{P}\left[\left|\deg_G(I) - 2^{\binom{|I|}{2}-\binom{2d}{2}}\binom{n-|I|}{2d-|I|} - \left(|A_I| - \frac{n-|I|}{2^{|I|}}\right)\frac{2^{\binom{|I|+1}{2}-\binom{2d}{2}}(n-|I|)^{2d-|I|-1}}{(2d-|I|-1)!}\right| > C2^{2d}(3-\ln\varepsilon)^2 n^{2d-|I|-1}\right] < \varepsilon.$$

The key lemma in proving Theorem 4.7.2 is the following, which bounds how often subsets of G of size x share (potential) edges.

Lemma 4.7.5. *If $x \geqslant 2$ and $n \geqslant x^2q(q+2)$ then there are at most $2n^{xq-q}\left(\frac{x^2}{x!}\right)^q$ multi-sets $S = \{V_1, \cdots, V_q\}$ of subsets $V_i \subseteq [n]$ of size $|V_i| = x$ such that that for all j there exists an $i \neq j$ such that $|V_i \cap V_j| \geqslant 2$.*

Using Lemma 4.7.5, we can bound the deviation of the number of x-cliques in G from its expected value; we carry this out now.

Definition 4.7.6. Define $X = \sum_{V:V\subseteq G,|V|=x}\left(1_V - 2^{-\binom{x}{2}}\right)$ where $1_V = 1$ if V is a clique in G and 0 otherwise.

Proposition 4.7.7. $X = N_x(G) - 2^{-\binom{x}{2}}\binom{n}{x}$

Proof. By observation. □

Corollary 4.7.8. *If $x \geqslant 2$ and $n \geqslant x^2q(q+2)$ then $E[X^q] < 2q!\left(\frac{x^2}{x!}n^{x-1}\right)^q$.*

Proof. By Proposition 4.7.7, $\mathbb{E}[X^q] = \sum_{\substack{V_1,\cdots,V_q:\\ \forall i, V_i\subseteq V(G), |V_i|=x}} \mathbb{E}\left[\prod_{i=1}^q \left(1_{V_i} - 2^{-\binom{x}{2}}\right)\right]$. Note that all terms of this sum have value less than 1. Furthermore, for all nonzero terms in this sum, for all j there must be an i such that $|V_i \cap V_j| \geqslant 2$, since the sets V_i and V_j must share a potential edge in order for 1_{V_i} and 1_{V_j} not to be independent. Thus, this sum is at most the number of ordered multi-sets of q x-cliques $\{V_1, \cdots, V_q\}$ where for all j there is an i such that $|V_i \cap V_j| \geqslant 2$. In turn, this is at most $q!$ times the number of unordered multi-sets of such q x-cliques. By Lemma 4.7.5, this is at most $2q!\left(\frac{x^2}{x!}n^{x-1}\right)^q$, as needed. □

We are now ready to prove Theorem 4.7.2.

Proof of Theorem 4.7.2. The result is trivial for $x = 0$ and $x = 1$ so we may assume that $x \geqslant 2$. Using Corollary 4.7.8 and Markov's inequality,

$$\varepsilon > \mathbb{P}\left[X^q > \frac{E[X^q]}{\varepsilon}\right] \geqslant \mathbb{P}\left[X^q > \frac{2q!\left(\frac{x^2}{x!}n^{x-1}\right)^q}{\varepsilon}\right] = \mathbb{P}\left[|X| > \sqrt[q]{\frac{2q!}{\varepsilon}}\frac{x^2}{x!}n^{x-1}\right]$$

Thus, we just need to give an upper bound on $\min_{\{\text{positive even q}\}}\{\sqrt[q]{\frac{2q!}{\varepsilon}}\}$. For all positive even q, $2q! \leqslant q^q$ so this expression is upper bounded by $\frac{q}{\sqrt[q]{\varepsilon}}$. We now try to minimize $\frac{q}{\sqrt[q]{\varepsilon}}$ over all positive even q. Taking the derivative of this expression with respect to q yields $\frac{1}{\sqrt[q]{\varepsilon}} + \frac{\ln \varepsilon}{q\sqrt[q]{\varepsilon}}$. Setting this to 0 yields $q = -\ln \varepsilon$. However, we require q to be even so we take q to be the smallest positive even integer which is greater than $-\ln \varepsilon$. Now $q < 2 - \ln \varepsilon$ and $\sqrt[q]{\varepsilon} \geqslant \sqrt[(-\ln \varepsilon)]{\varepsilon} = \left(e^{\ln \varepsilon}\right)^{\frac{1}{-\ln \varepsilon}} = \frac{1}{e}$. Putting everything together, for this q, $\sqrt[q]{\frac{2q!}{\varepsilon}} \leqslant \frac{q}{\sqrt[q]{\varepsilon}} < 2e - e\ln \varepsilon$. Plugging this in gives

$$\varepsilon > P\left[|X| > \sqrt[q]{\frac{2q!}{\varepsilon}}\frac{x^2}{x!}n^{x-1}\right] \geqslant P\left[|X| > e(2 - \ln \varepsilon)\frac{x^2}{x!}n^{x-1}\right]$$

All that is left is to check that $n \geqslant x^2 q(q+2)$ for this q to make sure that our application of Corollary 4.7.8 was valid. Since $n > x^2(2e - e\ln \varepsilon)(2e + 2 - e\ln \varepsilon)$, this holds, as needed. □

Now that we have proven Theorem 4.7.2, we will derive Theorem 4.7.4 from Theorem 4.7.2. The idea is that conditioned on I being a clique, by Theorem 4.7.2, $\deg_G(I)$ is primarily determined by $|A_I|$, which can be easily shown to be tightly concentrated around its expected value. We start with the following lemma

Lemma 4.7.9. *If $n > d$ then for any $I \subseteq [n]$ of size less than d, if we first determine all of the edges incident to elements of I (which determines A_I) then if I is a clique, when we look at the remainder*

of the graph, for any $\varepsilon_1 \in (0,1)$,

$$P\Big(\left|\deg_G(I) - 2^{\binom{|I|}{2}-\binom{d}{2}}\binom{n-|I|}{d-|I|} - \left(|A_I| - \frac{n-|I|}{2^{|I|}}\right)\frac{2^{\binom{|I|+1}{2}-\binom{d}{2}}(n-|I|)^{d-|I|-1}}{(d-|I|-1)!}\right| >$$
$$10(2-\ln\varepsilon_1)n^{d-|I|-1} + (d-|I|)^2\left(\frac{2^{|I|}|A_I|}{n-|I|}-1\right)^2\frac{2^{\binom{|I|}{2}-\binom{d}{2}}(n-|I|)^{d-|I|}}{(d-|I|)!}\Big) < \varepsilon_1$$

so long as the following conditions hold:

1. $(d-|I|)\left|\frac{2^{|I|}|A_I|}{n-|I|}-1\right| \leqslant 1$
2. $|A_I| > d^2(2e - e\ln\varepsilon_1)(2e+2-e\ln\varepsilon_1)$

To prove Lemma 4.7.9 we require the following results; proofs of the more elementary ones are deferred to Section 4.7.1.

Proposition 4.7.10. *For all nonnegative integers* n *and* k *where* $k < n$, $0 \leqslant \frac{n^k}{k!} - \binom{n}{k} \leqslant \frac{k^2}{2n}\frac{n^k}{k!}$

Proof of Proposition 4.7.10. Note that $n^k \geqslant \prod_{j=0}^{k-1}(n-j) \geqslant n^k\prod_{j=0}^{k-1}(1-\frac{j}{n}) \geqslant n^k(1-\sum_{j=0}^{k-1}\frac{j}{n}) \geqslant n^k(1-\frac{k^2}{2n})$

This implies that $0 \leqslant n^k - \prod_{j=0}^{k-1}(n-j) \leqslant \frac{k^2}{2n}n^k$ and dividing everything by $k!$ gives the claimed result. □

Lemma 4.7.11. *If* $n > d \geqslant |I|$ *then* $\left|2^{\binom{|I|}{2}-\binom{d}{2}}\frac{(n-|I|)^{d-|I|}}{(d-|I|)!} - 2^{\binom{|I|}{2}-\binom{d}{2}}\binom{n-|I|}{d-|I|}\right| \leqslant 2^{\binom{|I|}{2}-\binom{d}{2}}n^{d-|I|-1}$

Proof of Lemma 4.7.11. Applying Proposition 4.7.10 on $n-|I|$ and $d-|I|$ gives

$$\left|\frac{(n-|I|)^{d-|I|}}{(d-|I|)!} - \binom{n-|I|}{d-|I|}\right| \leqslant \frac{(d-|I|)^2}{2(n-|I|)}\frac{(n-|I|)^{d-|I|}}{(d-|I|)!} \leqslant (n-|I|)^{d-|I|-1} \leqslant n^{d-|I|-1}$$

Multiplying this equation by $2^{\binom{|I|}{2}-\binom{d}{2}}$ gives the claimed result. □

Proposition 4.7.12. *For all nonnegative integers x and d where $x \leqslant d$, $x(d-x)+\binom{x}{2}-\binom{d}{2} = -\binom{d-x}{2}$.*

Proposition 4.7.13. *For any nonnegative integer k and any x such that $|kx| \leqslant 1$, $\left|(1+x)^k - (1+kx)\right| \leqslant k^2x^2$*

Eventually in the course of proving Lemma 4.7.9 we will break

$$\deg_G(I) - 2^{\binom{|I|}{2}-\binom{d}{2}}\binom{n-|I|}{d-|I|} - \left(|A_I| - \frac{n-|I|}{2^{|I|}}\right)\frac{2^{\binom{|I|+1}{2}-\binom{d}{2}}(n-|I|)^{d-|I|-1}}{(d-|I|-1)!}$$

into pieces. The following lemmas offer the necessary bounds on each piece.

Lemma 4.7.14. *If $(d-|I|)\left|\frac{2^{|I|}|A_I|}{n-|I|}-1\right| \leqslant 1$ then*

$$\left|2^{-\binom{d-|I|}{2}}\frac{|A_I|^{d-|I|}}{(d-|I|)!} - \frac{2^{\binom{|I|}{2}-\binom{d}{2}}(n-|I|)^{d-|I|}}{(d-|I|)!} - \left(|A_I| - \frac{n-|I|}{2^{|I|}}\right)\frac{2^{\binom{|I|+1}{2}-\binom{d}{2}}(n-|I|)^{d-|I|-1}}{(d-|I|-1)!}\right|$$
$$\leqslant (d-|I|)^2\left(\frac{2^{|I|}|A_I|}{n-|I|}-1\right)^2\frac{2^{\binom{|I|}{2}-\binom{d}{2}}(n-|I|)^{d-|I|}}{(d-|I|)!}$$

Proof of Lemma 4.7.14. Applying Proposition 4.7.13 with $x = \frac{2^{|I|}|A_I|}{n-|I|} - 1$ and $k = (d-|I|)$, since $(d-|I|)\left|\frac{2^{|I|}|A_I|}{n-|I|}-1\right| \leqslant 1$,

$$\left|\frac{2^{|I|(d-|I|)}|A_I|^{d-|I|}}{(n-|I|)^{d-|I|}} - 1 - (d-|I|)\left(\frac{2^{|I|}|A_I|}{n-|I|}-1\right)\right| \leqslant (d-|I|)^2\left(\frac{2^{|I|}|A_I|}{n-|I|}-1\right)^2$$

Multiplying this equation by $\frac{2^{\binom{|I|}{2}-\binom{d}{2}}(n-|I|)^{d-|I|}}{(d-|I|)!}$ and using Proposition 4.7.12 with $x = |I|$ gives

$$\left|2^{-\binom{d-|I|}{2}}\frac{|A_I|^{d-|I|}}{(d-|I|)!} - \frac{2^{\binom{|I|}{2}-\binom{d}{2}}(n-|I|)^{d-|I|}}{(d-|I|)!} - (d-|I|)\left(\frac{2^{|I|}|A_I|}{n-|I|}-1\right)\frac{2^{\binom{|I|}{2}-\binom{d}{2}}(n-|I|)^{d-|I|}}{(d-|I|)!}\right|$$
$$= \left|2^{-\binom{d-|I|}{2}}\frac{|A_I|^{d-|I|}}{(d-|I|)!} - \frac{2^{\binom{|I|}{2}-\binom{d}{2}}(n-|I|)^{d-|I|}}{(d-|I|)!} - \left(A_I - \frac{n-|I|}{2^{|I|}}\right)\frac{2^{\binom{|I|+1}{2}-\binom{d}{2}}(n-|I|)^{d-|I|-1}}{(d-|I|-1)!}\right|$$

$$\leqslant (d-|I|)^2 \left(\frac{2^{|I|}A_I}{n-|I|} - 1\right)^2 \frac{2^{\binom{|I|}{2}-\binom{d}{2}}(n-|I|)^{d-|I|}}{(d-|I|)!}$$

□

Lemma 4.7.15. *If $|A_I| > d - |I|$ then* $\left|2^{-\binom{d-|I|}{2}}\binom{|A_I|}{d-|I|} - 2^{-\binom{d-|I|}{2}}\frac{|A_I|^{d-|I|}}{(d-|I|)!}\right| \leqslant 2^{-\binom{d-|I|}{2}}|A_I|^{d-|I|-1}$

Proof of Lemma 4.7.15. Applying Proposition 4.7.10 on $|A_I|$ and $d - |I|$ gives

$$\left|\frac{|A_I|^{d-|I|}}{(d-|I|)!} - \binom{|A_I|}{d-|I|}\right| \leqslant \frac{(d-|I|)^2}{2|A_I|}\frac{|A_I|^{d-|I|}}{(d-|I|)!} \leqslant |A_I|^{d-|I|-1}$$

Multiplying this equation by $2^{-\binom{d-|I|}{2}}$ gives the claimed result. □

Lemma 4.7.16. *For all $\varepsilon_1 \in (0,1)$, if $|A_I| > d^2(2e - e\ln\varepsilon_1)(2e + 2 - e\ln\varepsilon_1)$ then*

$$\mathbb{P}\left[\left|\deg_G(I) - 2^{-\binom{d-|I|}{2}}\binom{|A_I|}{d-|I|}\right| > e(2-\ln\varepsilon_1)\frac{(d-|I|)^2}{(d-|I|)!}|A_I|^{d-|I|-1}\right] < \varepsilon_1$$

Proof of Lemma 4.7.16. This lemma follows immediately from applying Theorem 4.7.2 on the random graph G restricted to the vertices A_I. □

Proof of Lemma 4.7.9. We break

$$\deg_G(I) - 2^{\binom{|I|}{2}-\binom{d}{2}}\binom{n-|I|}{d-|I|} - \left(|A_I| - \frac{n-|I|}{2^{|I|}}\right)\frac{2^{\binom{|I|+1}{2}-\binom{d}{2}}(n-|I|)^{d-|I|-1}}{(d-|I|-1)!}$$

into four parts and analyze each one separately.

1. $2^{\binom{|I|}{2}-\binom{d}{2}}\frac{(n-|I|)^{d-|I|}}{(d-|I|)!} - 2^{\binom{|I|}{2}-\binom{d}{2}}\binom{n-|I|}{d-|I|}$
2. $2^{-\binom{d-|I|}{2}}\frac{|A_I|^{d-|I|}}{(d-|I|)!} - 2^{\binom{|I|}{2}-\binom{d}{2}}\frac{(n-|I|)^{d-|I|}}{(d-|I|)!} - \left(|A_I| - \frac{n-|I|}{2^{|I|}}\right)\frac{2^{\binom{|I|+1}{2}-\binom{d}{2}}(n-|I|)^{d-|I|-1}}{(d-|I|-1)!}$
3. $2^{-\binom{d-|I|}{2}}\binom{|A_I|}{d-|I|} - 2^{-\binom{d-|I|}{2}}\frac{|A_I|^{d-|I|}}{(d-|I|)!}$

4. $\deg_G(I) - 2^{-\binom{d-|I|}{2}}\binom{|A_I|}{d-|I|}$

Combining Lemma 4.7.11, Lemma 4.7.14, Lemma 4.7.15, and Lemma 4.7.16, we have that under the given conditions,

$$\mathbb{P}\left(\left|\deg_G(I) - 2^{\binom{|I|}{2}-\binom{d}{2}}\binom{n-|I|}{d-|I|} - \left(|A_I| - \frac{n-|I|}{2^{|I|}}\right)\frac{2^{\binom{|I|+1}{2}-\binom{d}{2}}(n-|I|)^{d-|I|-1}}{(d-|I|-1)!}\right| > \right.$$
$$2^{\binom{|I|}{2}-\binom{d}{2}}n^{d-|I|-1} + (d-|I|)^2\left(\frac{2^{|I|}|A_I|}{n-|I|} - 1\right)^2\frac{2^{\binom{|I|}{2}-\binom{d}{2}}(n-|I|)^{d-|I|}}{(d-|I|)!} +$$
$$\left. 2^{-\binom{d-|I|}{2}}|A_I|^{d-|I|-1} + e(2-\ln\varepsilon_1)\frac{(d-|I|)^2}{(d-|I|)!}|A_I|^{d-|I|-1}\right) < \varepsilon_1$$

The result now reduces to showing the following equation

$$2^{\binom{|I|}{2}-\binom{d}{2}}n^{d-|I|-1} + 2^{-\binom{d-|I|}{2}}|A_I|^{d-|I|-1} + e(2-\ln\varepsilon_1)\frac{(d-|I|)^2}{(d-|I|)!}|A_I|^{d-|I|-1} \leqslant 10(2-\ln\varepsilon_1)n^{d-|I|-1}$$

which follows from the facts that $|I| < d$, $|A_I| \leqslant n$, and $\frac{(d-|I|)^2}{(d-|I|)!} \leqslant 2$. □

To use Lemma 4.7.9 to prove Theorem 4.7.4, we need probabilistic bounds on $|A_I|$.

Lemma 4.7.17. *There is a universal constant C so that for all $\varepsilon_2 \in (0,1)$, $P\left[||A_I| - 2^{-|I|}(n-|I|)| > C(2-\ln\varepsilon_2)\sqrt{n}\right] < \varepsilon_2$*

Proof. The lemma follows from standard concentration of measure. If we let x_i be 1 if $i \notin I$ and i is adjancent to all vertices in I and 0 otherwise then $\sum_{i=1}^n x_i = A_I$. The expected value of $A_I = \sum_{i=1}^n x_i$ is $2^{-|I|}(n-|I|)$, so by Bernstein's inequality there is C so that for all $\varepsilon_2 \in (0,1)$, $\mathbb{P}\left[|A_I - 2^{-|I|}(n-|I|)| > C(2-\ln\varepsilon_2)\sqrt{n}\right] < \varepsilon_2$. □

We have all we need now to prove Theorem 4.7.4.

Proof of Theorem 4.7.4. The result is trivial if $|I| = d$ and follows immediately from Theorem 4.7.2 if $|I| = 0$ so we may assume that $0 < |I| < d$.

In what follows, C and C' denotes universal constants which may vary from line to line. Now recall that by Lemma 4.7.9, for any $\varepsilon_1 \in (0,1)$,

$$P\Big(\left|\deg_G(I) - 2^{\binom{|I|}{2}-\binom{d}{2}}\binom{n-|I|}{d-|I|} - \left(|A_I| - \frac{n-|I|}{2^{|I|}}\right)\frac{2^{\binom{|I|+1}{2}-\binom{d}{2}}(n-|I|)^{d-|I|-1}}{(d-|I|-1)!}\right| >$$
$$C(2-\ln\varepsilon_1)n^{d-|I|-1} + (d-|I|)^2\left(\frac{2^{|I|}|A_I|}{n-|I|} - 1\right)^2\frac{2^{\binom{|I|}{2}-\binom{d}{2}}(n-|I|)^{d-|I|}}{(d-|I|)!}\Big) < \varepsilon_1$$

so long as the following conditions hold:

1. $(d-|I|)\left|\frac{2^{|I|}|A_I|}{n-|I|} - 1\right| \leqslant 1$
2. $|A_I| > d^2(2e - e\ln\varepsilon_1)(2e+2-e\ln\varepsilon_1)$

Taking $\varepsilon_1 = \varepsilon_2 = \frac{\varepsilon}{2}$, plugging Lemma 4.7.17 into these equations and using the union bound, we have that

$$P\Big(\left|\deg_G(I) - 2^{\binom{|I|}{2}-\binom{d}{2}}\binom{n-|I|}{d-|I|} - \left(|A_I| - \frac{n-|I|}{2^{|I|}}\right)\frac{2^{\binom{|I|+1}{2}-\binom{d}{2}}(n-|I|)^{d-|I|-1}}{(d-|I|-1)!}\right| >$$
$$C(3-\ln\varepsilon)n^{d-|I|-1} + (d-|I|)^2\left(\frac{2^{|I|}C'(3-\ln\varepsilon)\sqrt{n}}{n-|I|}\right)^2\frac{2^{\binom{|I|}{2}-\binom{d}{2}}(n-|I|)^{d-|I|}}{(d-|I|)!}\Big) < \varepsilon$$

so long as the corresponding conditions hold. Assuming these conditions hold for now, since $|I| < \frac{n}{16}$, $|I| < d$, $\frac{(d-|I|)^2}{(d-|I|)!} \leqslant 2$, and $2^{|I|}2^{\binom{|I|}{2}} = 2^{\binom{|I|+1}{2}}$,

$$(d-|I|)^2\left(\frac{2^{|I|}C(3-\ln\varepsilon)\sqrt{n}}{n-|I|}\right)^2\frac{2^{\binom{|I|}{2}-\binom{d}{2}}(n-|I|)^{d-|I|}}{(d-|I|)!} \leqslant C'\frac{2^{|I|}(3-\ln\varepsilon)^2 n}{(n-|I|)}(n-|I|)^{d-|I|-1}$$
$$< C\cdot 2^d(3-\ln\varepsilon)^2 n^{d-|I|-1}$$

Plugging this in we have that

$$P\Big(\Big|\deg_G(I) - 2^{\binom{|I|}{2}-\binom{d}{2}}\binom{n-|I|}{d-|I|} - \left(|A_I| - \frac{n-|I|}{2^{|I|}}\right)\frac{2^{\binom{|I|+1}{2}-\binom{d}{2}}(n-|I|)^{d-|I|-1}}{(d-|I|-1)!}\Big| > C2^d(3-\ln\varepsilon)^2 n^{d-|I|-1}\Big) < \varepsilon$$

as needed. For the first part of Theorem Theorem 4.7.4, note that this implies that

$$P\Big(\Big|\deg_G(I) - 2^{\binom{|I|}{2}-\binom{d}{2}}\binom{n-|I|}{d-|I|}\Big| > \Big||A_I| - \frac{n-|I|}{2^{|I|}}\Big|\frac{2^{\binom{|I|+1}{2}-\binom{d}{2}}(n-|I|)^{d-|I|-1}}{(d-|I|-1)!} + C2^d(3-\ln\varepsilon)^2 n^{d-|I|-1}\Big) < \varepsilon$$

Plugging in Lemma 4.7.17 and noting that

$$C(3-\ln\varepsilon)\sqrt{n}\frac{2^{\binom{|I|+1}{2}-\binom{d}{2}}(n-|I|)^{d-|I|-1}}{(d-|I|-1)!} < C'(3-\ln\varepsilon)n^{d-|I|-\frac{1}{2}}$$

we have that

$$\mathbb{P}\left(\left|\deg_G(I) - 2^{\binom{|I|}{2}-\binom{d}{2}}\binom{n-|I|}{d-|I|}\right| > C(3-\ln\varepsilon)n^{d-|I|-\frac{1}{2}} + C'2^d(3-\ln\varepsilon)^2 n^{d-|I|-1}\right) < \varepsilon$$

Taking $n \geqslant Cd^2 2^{2d}(3-\ln\varepsilon)^2$ and $d \geqslant 2$,

$$C2^d(3-\ln\varepsilon)^2 n^{d-|I|-1} \leqslant C'(3-\ln\varepsilon)n^{d-|I|-\frac{1}{2}}$$

Plugging this in gives that

$$\mathbb{P}\left(\left|\deg_G(I) - 2^{\binom{|I|}{2}-\binom{d}{2}}\binom{n-|I|}{d-|I|}\right| > C(3-\ln\varepsilon)n^{d-|I|-\frac{1}{2}}\right) < \varepsilon$$

as needed. All that is left is to check the conditions for Lemma 4.7.9, which are as follows.

1. $(d-|I|)\frac{2^{|I|}e(3-\ln\varepsilon)\sqrt{n}}{n-|I|} \leqslant 1$

2. $2^{-|I|}(n-|I|) > d^2e(3-\ln\varepsilon)(3e+2-e\ln\varepsilon)+e(3-\ln\varepsilon)\sqrt{n}$

These conditions are true if $n \geqslant 4d^2 2^{2d}(3-\ln\varepsilon)^2$. To see this, note that since $d > |I| > 0$ and $|I| < \frac{n}{16}$

1. $(d-|I|)2^{|I|}e(3-\ln\varepsilon)\sqrt{n} \leqslant d2^{|I|}e(3-\ln\varepsilon)\sqrt{n}-|I| \leqslant \frac{e}{4}\sqrt{4d^2 2^{2d}(3-\ln\varepsilon)^2}\sqrt{n}-|I| \leqslant n-|I|$
2. $2^{|I|}d^2e(3-\ln\varepsilon)(3e+2-e\ln\varepsilon) < 2^{|I|}d^2e^2\frac{3e+2}{3e}(3-\ln\varepsilon)^2 < 10\cdot 2^{|I|}d^2(3-\ln\varepsilon)^2 \leqslant \frac{5}{16}n$
3. $2^{|I|}e(3-\ln\varepsilon)\sqrt{n} < 4\cdot 2^{|I|}(3-\ln\varepsilon)\left(2d2^d(3-\ln\varepsilon)\right) \leqslant \frac{n}{2}$

Dividing the first statement by $n-|I|$ gives the first condition. Using the second and third statements,

$$2^{|I|}d^2e(3-\ln\varepsilon)(3e+2-e\ln\varepsilon)+2^{|I|}e(3-\ln\varepsilon)\sqrt{n} < \frac{13}{16}n < n-|I|$$

Dividing this by $2^{|I|}$ gives the second condition, as needed. □

4.7.1 Proofs of Remaining Lemmas

Proof of Lemma 4.7.5.

Definition 4.7.18. For each multi-set $S = \{V_1, \cdots, V_k\}$ of k x-cliques, define the constraint graph H_S as follows.

1. $V(H_S) = \{V_1, \cdots, V_k\}$
2. $E(H_S) = \{(V_i, V_j) : |V_i \cap V_j| \geqslant 2\}$

Let's first bound the number of S such that H_S is connected.

Lemma 4.7.19. *For any $x, k \geqslant 2$, there are at most $n^{kx-2k+2}k!\frac{2}{x^4}\left(\frac{x^4}{2(x!)}\right)^k$ multi-sets S of k x-cliques such that H_S is connected.*

Proof. Since H_S is connected, we can order $\{V_1, \cdots, V_k\}$ so that for all $j > 1$ there is an $i < j$ such that $(V_i, V_j) \in E(H_S)$. Assuming this is the case, we have at most $\binom{n}{x}$ choices for V_1. For each $j > 1$, there are two vertices in V_j which are contained in some i where $i < j$. There are at most j choices for this i and then there are $\binom{d}{2}$ choices for which two vertices of V_i are contained in V_j. There are at most $\binom{n}{x-2}$ choices for the other $x-2$ vertices of V_j so for each $j > 1$ there are at most $j\binom{x}{2}\binom{n}{x-2}$ choices for V_j. Putting everything together, there are at most

$$\binom{n}{x}k!\left(\binom{x}{2}\binom{n}{x-2}\right)^{k-1} \leqslant n^x\frac{k!}{x!}\left(\frac{x^4n^{x-2}}{2(x!)}\right)^{k-1} \leqslant n^{kx-2k+2}k!\frac{2}{x^4}\left(\frac{x^4}{2(x!)}\right)^k$$

multi-sets S of k x-cliques such that H_S is connected. □

Now consider the number of multi-sets S of q x-cliques such that H_S has t connected components with sizes $s_1, \cdots, s_t$. Using Lemma 4.7.19, there are at most

$$\prod_{i=1}^{t}\left(n^{s_ix-2s_i+2}s_i!\frac{2}{x^4}\left(\frac{x^4}{2(x!)}\right)^{s_i}\right) = \left(\prod_{i=1}^{t}s_i!\right)n^{xq-2q+2t}\left(\frac{2}{x^4}\right)^t\left(\frac{x^4}{2(x!)}\right)^q$$

such S.

We now total this up over all possible $t, s_1, \cdots, s_t$. For the special case that all connected components of H_S have size 2, there are at most $n^{xq-q}\left(\frac{x^2}{x!}\right)^q$ such S. We will show that this term contributes more than all of the other terms combined, which implies that the total number of S is at most $2n^{xq-q}\left(\frac{x^2}{x!}\right)^q$, as needed.

For a given t, $\prod_{i=1}^{t} s_i! \leqslant 2^{t-1}(q+2-2t)! \leqslant 2^t(q+2)^{q-2t}$. Also, each of the t components of H_S must have at least two vertices so to determine the sizes $s_1, \cdots, s_t$ it is sufficient to decide how to distribute the $q-2t$ extra vertices among the t connected components of H_S. There are at most $\binom{q-t-1}{t-1} \leqslant q^{q-2t}$ ways to do this. Thus, the total contribution for terms of a given t is at most

$$2^t(q(q+2))^{q-2t} n^{xq-2q+2t} \left(\frac{2}{x^4}\right)^t \left(\frac{x^4}{2(x!)}\right)^q = \left(\frac{x^2 q(q+2)}{2n}\right)^{q-2t} \left(n^{xq-q}\left(\frac{x^2}{(x!)}\right)^q\right)$$

Since $n \geqslant x^2 q(q+2)$, the $n^{xq-q}\left(\frac{x^2}{x!}\right)^q$ term which comes from $t = \frac{q}{2}$ contributes more than all the other terms combined, as needed. □

Proof of Proposition 4.7.12. Rearranging this equation gives $\binom{d}{2} = \binom{x}{2} + dx + \binom{d-x}{2}$ which just says that if we want to pick two elements in $[1, d]$ we can either pick two elements in $[1, x]$, one element from $[1, x]$ and one element from $[x+1, d]$, or two elements from $[x+1, d]$. □

Proof of Proposition 4.7.13.

$$\left|(1+x)^k - (1+kx)\right| \leqslant \sum_{j=2}^{k} \left|\binom{k}{j} x^j\right| \leqslant \sum_{j=2}^{k} \left|\frac{1}{j!} k^j x^j\right| \leqslant k^2 x^2 \sum_{j=2}^{k} \frac{1}{j!} \leqslant k^2 x^2$$

□

4.8 Optimality of MPW Analysis

In this section we sketch an argument due to Kelner showing that the MPW moments are not PSD when $\omega \gg n^{1/(d+1)}$.

Theorem 4.8.1. *With high probability, the MW moments are not PSD when $\omega \gg n^{\frac{1}{d+1}}$. In particular, if $\omega \gg n^{\frac{1}{d+1}}$ then for all s, for some appropriately chosen C, with high probability,*

$$\tilde{E}[(C\omega^d x_s - \sum_{I:I\subseteq V,|I|=d} \prod_{i\in I} r_s(i)x_i)^2] < 0$$

with high probability.

Proof Sketch. We will be using the following proposition heavily.

Proposition 4.8.2. *For all $I \subseteq V(G)$ such that $|I| < 2d$, $\tilde{E}[\sum_{j\notin I} x_{I\cup j}] = (\omega - |I|)\tilde{E}[x_I]$*

Proof. We have the equation that $\sum_j x_j = \omega$ so

$$\tilde{E}[\sum_j x_{I\cup j}] = |I|\tilde{E}[x_I] + \tilde{E}[\sum_{j\notin I} x_{I\cup j}] = \omega\tilde{E}[x_I]$$

and the result follows. □

Corollary 4.8.3. *For all $I \subseteq V(G)$ and all m such that $|I| + m \leqslant 2d$,*

$$\tilde{E}[\sum_{J:I\subseteq J,|J|=|I|+m} x_J] = \frac{\prod_{x=0}^{m-1}(\omega - |I| - x)}{m!}\tilde{E}[x_I] = \binom{\omega - |I|}{m}\tilde{E}[x_I]$$

Proof. This result follows from repeatedly expanding out $(\omega - |K|)\tilde{E}[x_K] = \sum_{j\notin K} \tilde{E}[x_{K\cup j}]$. For any given J such that $I \subseteq J$ and $|J| = |I| + m$ there are $m!$ different ways to reach J from I which gives us the $m!$ term. □

With this corollary in hand, we expand out $\tilde{E}[(C\omega^d x_s - \sum_{\substack{I:I\subseteq V\setminus\{s\},\\ |I|=d}} \prod_{i\in I} r_s(i)x_i)^2]$, obtaining

$$\tilde{E}[C^2\omega^{2d}x_s] - 2C\omega^d\tilde{E}[\sum_{\substack{I:I\subseteq V\setminus\{s\},\\ |I|=d}} x_{I\cup\{s\}}] + \tilde{E}[\sum_{\substack{I,J:I,J\subseteq V\setminus\{s\},\\ |I|=|J|=d}} \left(\prod_{i\in I\Delta J} r_s(i)\right) x_{I\cup J}]$$

Using Corollary 4.8.3,

$$2C\omega^d \tilde{E}[\sum_{\substack{I:I\subseteq V\setminus\{s\},\\ |I|=d}} x_{I\cup\{s\}}] = 2C\omega^d \binom{\omega-1}{d} \tilde{E}[x_s]$$

Thus combining the first two terms gives

$$\left(C^2\omega^{2d} - 2C\omega^d \binom{\omega-1}{d}\right) \tilde{E}[x_s]$$

From our concnetration bounds on $deg_G(s)$, with high probability $\tilde{E}(x_s)$ is $\frac{\omega}{n}(1 \pm O(\frac{log(n)}{\sqrt{n}}))$. Thus, taking C to be a sufficiently small constant (which will depend on d), with high probability the first two terms are $-\Omega(\frac{\omega^{2d+1}}{n})$

Analyzing the third term is trickier. For the third term, taking $K = I\Delta J$ for each term,

$$\begin{aligned}\tilde{E}[\sum_{\substack{I,J:I,J\subseteq V\setminus\{s\},\\ |I|=|J|=d}} \left(\prod_{i\in I\Delta J} r_s(i)\right) x_{I\cup J}] &= \sum_{x=0}^{d} \sum_{\substack{I,J:I,J\subseteq V\setminus\{s\},\\ |I|=|J|=d, |I\Delta J|=2x}} \left(\prod_{i\in I\Delta J} r_s(i)\right) \tilde{E}[x_{I\cup J}] \\ &= \sum_{x=0}^{d} \sum_{\substack{K:K\subseteq V\setminus\{s\},\\ |K|=2x}} \binom{\omega-d-x}{d-x} \left(\prod_{i\in K} r_s(i)\right) \tilde{E}[x_K]\end{aligned}$$

We will now analyze the expected value and variance of this expression. However, before doing so there is a subtle issue we must deal with. $\tilde{E}[x_K]$ is not completely independent of $(\prod_{i\in K} r_s(i))$. What saves us is that the dependence is small enough to be negligible.

For each $K \subseteq V(G)\setminus\{s\}$, define y_K to be the expected value of $\tilde{E}[x_K]$ if we preserve all of the edges of G which are not incident with s but reselect the edges of G incident

with s randomly. From our concentration results on $deg_G(K)$, with high probability, for all K, $|\tilde{E}[x_K] - y_K|$ is $O(\frac{\omega^{|K|}log(n)}{n^{|K|+1}})$. We now write

$$\sum_{x=0}^{d} \sum_{\substack{K:K\subseteq V\setminus\{s\},\\ |K|=2x}} \binom{\omega-d-x}{d-x}\left(\prod_{i\in K} r_s(i)\right)\tilde{E}[x_K] =$$

$$\sum_{x=0}^{d} \sum_{\substack{K:K\subseteq V\setminus\{s\},\\ |K|=2x}} \binom{\omega-d-x}{d-x}\left(\prod_{i\in K} r_s(i)\right) y_K$$

$$+ \sum_{x=0}^{d} \sum_{\substack{K:K\subseteq V\setminus\{s\},\\ |K|=2x}} \binom{\omega-d-x}{d-x}\left(\prod_{i\in K} r_s(i)\right)(\tilde{E}[x_K] - y_K)$$

For the second part, for each $x \in [0, d]$ there are at most $\binom{n}{2x}$ K such that $K \subseteq V(G) \setminus \{s\}$ and $|K| = 2x$. Thus,

$$\sum_{x=0}^{d} \sum_{\substack{K:K\subseteq V\setminus\{s\},\\ |K|=2x}} \binom{\omega-d-x}{d-x}\left(\prod_{i\in K} r_s(i)\right)(\tilde{E}[x_K] - y_K)$$

$$\leqslant \sum_{x=0}^{d} \binom{n}{2x}\binom{\omega-d-x}{d-x} \max_{K:|K|=2x} |\tilde{E}[x_K] - y_K|$$

From our concentration bounds, with high probability, for all x the corresponding term on the right is $O(\frac{\omega^{d+x}\lg n}{n})$ which is $O(\frac{\omega^{d+x}}{n})$. For all $x \leqslant d$, this is much smaller than $\Omega(\frac{\omega^{2d+1}}{n})$. Thus we may ignore the second part.

For the first part, the values $r_s(i)$ are completely independent of the values y_K. When we take the expected value of

$$\sum_{x=0}^{d} \sum_{\substack{K:K\subseteq V\setminus\{s\},\\ |K|=2x}} \binom{\omega-d-x}{d-x}\left(\prod_{i\in K} r_s(i)\right) y_K$$

over the edges incident to s, only the $x = 0$ term remains and we obtain $\binom{\omega-d}{d}$. When we take the variance of

$$\sum_{x=0}^{d} \sum_{\substack{K:K\subseteq V\setminus\{s\},\\ |K|=2x}} \binom{\omega-d-x}{d-x} \left(\prod_{i\in K} r_s(i)\right) y_K$$

over the edges incident to s, only the square terms remain so we obtain

$$\sum_{x=0}^{d} \sum_{\substack{K:K\subseteq V\setminus\{s\},\\ |K|=2x}} \binom{\omega-d-x}{d-x}^2 (y_K)^2$$

From our concentration bounds on the y_K, with high probability this is

$$\binom{\omega-d}{d}^2 + \sum_{x=1}^{d} \binom{\omega-d-x}{d-x}^2 \binom{n-1}{2x} O\left(\frac{\omega^{4x}}{n^{4x}}\right)$$

which is

$$\binom{\omega-d}{d}^2 + O\left(\frac{\omega^{2d+2}}{n^2}\right)$$

Thus, with high probability

$$\sum_{x=0}^{d} \sum_{\substack{K:K\subseteq V\setminus\{s\},\\ |K|=2x}} \binom{\omega-d-x}{d-x} \left(\prod_{i\in K} r_s(i)\right) y_K$$

is $\binom{\omega-d}{d}\left(1 \pm O(\frac{\omega}{n})\right)$ which is $O(\omega^d)$. Putting everything together, if $\omega \gg n^{\frac{1}{d+1}}$ then for some appropriately chosen C, with high probabbility, $\tilde{E}[(C\omega^d x_s - \sum_{I:I\subseteq V,|I|=d} \prod_{i\in I} r_s(i)x_i)^2] < 0$ □

Chapter 5

Tight Sum of Squares Lower Bound for Planted Clique

In this chapter[1], we present tight lower bound for the planted clique problem for the sum-of-squares algorithm at all degrees. The results of this chapter were obtained in joint work with Boaz Barak, Samuel Hopkins, Jonathan Kelner, Ankur Moitra and Aaron Potechin [21]. Formally, the main result of this chapter is the following theorem:

Theorem 5.0.1 (Optimal Planted Clique Lower Bound)**.** *There is an absolute constant c so that for every $d = d(n)$ and large enough n, the SOS relaxation of the planted clique problem has integrality gap at least $n^{1/2-c(d/\log n)^{1/2}}$.*

We begin by discussing the ways in which planted clique differs from problems for which strong SOS lower bounds have been shown before, and how this relates to a "computational Bayesian" perspective.

[1]The results of this chapter will appear in the Proceedings of Foundations of Computer Science, FOCS, 2016 in a paper titled *A Nearly Tight Sum-of-Squares Lower Bound for the Planted Clique Problem* co-authored with Boaz Barak, Samuel Hopkins, Jonathan Kelner, Ankur Moitra and Aaron Potechin. All the authors contributed equally in producing all the results in the paper and are listed in alphabetical order.

5.1 Planted Clique and Probabilistic Inference

There have been several strong lower bounds for the SOS algorithm before, in particular for problems such as 3SAT, 3XOR and other constraint satisfaction problems as well as the knapsack problem [59, 96, 20]. However, obtaining strong lower bounds for the planted clique problem seems to have required different techniques. A high-level way to describe the difference is that lower bounds for planted clique require accounting for **weak global constraints** rather than **strong local** ones. In the random 3SAT/3XOR setting, the effect of one variable on another is either extremely strong (if they are "nearby" in the formula) or essentially zero. In contrast in planted clique each variable has a weak *global* effect on all of the other variables. We now explain this in more detail.

Consider a random graph G in which a clique S of size ω has been planted. If someone tells us some simple statistics of G and then tells us that vertex 17 is not in S, this new informatoin makes it slightly less likely that 17's neighbors are in S and slightly more likely that 17's non-neighbors are in S. So, this information has a *weak global* effect. In contrast, when we have a random sparse 3SAT formula φ in which an assignment x has been planted, if someone tells us that $x_{17} = 0$ then it gives us a lot of information about the local neighborhood of the 17^{th} variable (the variables that are involved in constraints with 17 or one that have a short path of constraints to it) but there is an exponential decay of these correlations and so this information tells us almost nothing about the distribution of most of the variables x_i (that are far away from 17 in the sparse graph induced by φ)[2]. Thus, in the random 3SAT setting information about the assignments of individual

[2]This exponential decay can be shown formally for the case of satisfiable random 3SAT or 3XOR formulas whose clause density is sufficiently smaller than the threshold. In our regime of overconstrainted random

variables has a *strong local effect*. Indeed, previous Sum-of-Squares lower bounds for random 3SAT and 3XOR [59, 96] could be interpreted as producing a "distribution-like" object in which, conditioned on the value of a small set of variables S, some of the variables "close" to S in the formula were fixed, and the rest were completely independent.

This difference between the random SAT and the planted clique problems means that some subtleties that can be ignored in setting of random constraint satisfaction problems need to be tackled head-on when dealing with planted cliques. However to make this clearer, we need to take a detour and discuss Bayesian probabilities and their relation to Sum-of-Squares.

5.1.1 Computational Bayesian Probabilities and Pseudo-distributions

Strictly speaking, if a graph G contains a unique clique S of size ω, then for every vertex i the probability that i is in S is either zero or one. But, a computationally bounded observer may not know whether i is in the clique or not, and we could try to quantify this ignorance using probabilities. These can be thought of as a computational analogue of *Bayesian probabilities*, that, rather than aiming to measure the frequency at which an event occurs in some sample space, attempts to capture the subjective beliefs of some observer.

That is, the Bayesian probability that an observer B assigns to an event E can be thought of as corresponding to the odds at which B would make the bet that E holds. Note that this probability could be strictly between zero and one even if the event E is fully

3SAT/3XOR formulas there will not exist any satisfying assignments, and so to talk about "correlations" in the distributions of assignments we need to talk about the "Bayesian estimates" that arise from algorithms such as Sum-of-Squares or belief propagation. Both these algorithms exhibit this sort of exponential decay we talk about; see also Remark 5.1.1

determined, depending on the evidence available to B. While typically Bayesian analysis does not take into account computational limitations, one could imagine that even if B has access to information that fully determines whether E happened or not, he could still rationally assign a subjective probability to E that is strictly between zero and one if making the inferences from this information is computationally infeasible. In particular, in the example above, even if a computationally bounded observer has access to the graph G, which information-theoretically fully determines the planted ω-sized clique, he could still assign a probability strictly between zero and one to the event that vertex 17 is in the planted ω-sized clique, based on some simple to compute statistics such as how many neighbors 17 has, etc.

The Sum-of-Squares algorithm can be thought of as giving rise to an internally consistent set of such "computational probabilities". These probabilities may not capture *all* possible inferences that a computationally bounded observer could make, but they do capture all inferences that can be made via a powerful proof system.

Bayesian estimates for planted clique. To get a sense for our results and techniques, it is instructive to consider the following scenario. Let $G(n, 1/2, \omega)$ be the distribution over pairs (G, x) of an n-vertex graphs G and a vector $x \in \mathbb{R}^n$ which is obtained by sampling a random graph in $G(n, 1/2)$, planting an ω-sized clique in it, and letting G be the resulting graph and x the 0/1 characteristic vector of the planted clique. Let $f : \{0,1\}^{\binom{n}{2}} \times \mathbb{R}^n \to \mathbb{R}$ be some function that maps a graph G and a vector x into some real number $f_G(x)$. Now imagine two parties, Alice and Bob (where Bob can also stand for "Bayesian") that play the following game: Alice samples (G, x) from the distribution $G(n, 1/2, \omega)$ and sends G

to Bob, who wants to output the expected value of $f_G(x)$. We denote this value by $\tilde{\mathbb{E}}_G f_G$.

If we have no computational constraints then it is clear that Bob can simply output $\tilde{\mathbb{E}}_G f_G$ be equal to $\mathbb{E}_{x|G} f_G(x)$, by which we mean the expected value of $f_G(x)$ where x is chosen according to the conditional distribution on x given the graph G.[3] In particular, the value $\tilde{\mathbb{E}}_G f_G$ will be *calibrated* in the sense that

$$\mathop{\mathbb{E}}_{G \in_R G(n,1/2,\omega)} \tilde{\mathbb{E}}_G f_G = \mathop{\mathbb{E}}_{(G,x) \in_R G(n,1/2,\omega)} f_G(x) \tag{5.1.1}$$

Now if Bob is computationally bounded, then he will not necessarily be able to compute the value of $E_{x|G} f_G(x)$ even for a simple function such as $f_G(x) = x_{17}$. Indeed, as we mentioned, since with high probability the clique x is uniquely determined by G, $\mathbb{E}_{x|G} x_{17}$ will simply equal 1 if vertex 17 is in the clique and equal 0 otherwise. However, note that we don't need to compute the true conditional expectation to obtain a calibrated estimate. In the above example, if Bob simply outputs $\tilde{\mathbb{E}}\, x_{17} = \omega/n$ then his estimate will satisfy (5.1.1).

Our Sum-of-Squares lower bound amounts to coming up with some reasonable "pseudo-expectation" that can be efficiently computed, where $\tilde{\mathbb{E}}_G$ is meant to capture a "best effort" of a computationally bounded party of approximating the Bayesian conditional expectation $\mathbb{E}_{x|G}$. Our pseudo-expectation will be far from the true conditional expectations, but will be internally consistent in the sense that for all "simple" functions f it will satisfy (5.1.1). The key property is that our pseudo-expectation will not distinguish

[3] The astute reader might note that this expectation is somewhat degenerate since with very high probability the graph G will uniquely determine the vector x, but please bear with us, as in the computational setting we will be able to treat x as "undetermined".

between a graph G drawn from $G(n, 1/2, \omega)$ and a random G from $G(n, 1/2)$. In particular, it will also satisfy the following *pseudo-calibration* condition:

$$\mathop{\mathbb{E}}_{G \in_R G(n,1/2)} \tilde{\mathbb{E}}_G f_G = \mathop{\mathbb{E}}_{(G,x) \in_R G(n,1/2,\omega)} f_G(x) \tag{5.1.2}$$

for all "simple" functions $f = f(G, x)$. Note that (5.1.2) does not make sense for the estimates of a truly Bayesian (i.e., computationally unbounded) Bob, since almost all graphs G in $G(n, 1/2)$ are not even in the support of $G(n, 1/2, \omega)$. Nevertheless, our pseudo-distributions will be well defined even for a random graph and hence will yield estimates for the probabilities over this hypothetical object (i.e., the ω-sized clique) that does not exist. The "pseudo-calibration" condition (5.1.2) might seem innocent, but it turns out to imply many useful properties. In particular is not hard to see that (5.1.2) implies that for every *simple strong constraint* of the clique problem — a function f such that $f(G, x) = 0$ for every x that is a characteristic vector of an ω-clique in G — it must hold that $\tilde{\mathbb{E}}_G f_G = 0$. But even beyond these "strong constraints", (5.1.2) implies that the pseudo-expectation satisfies many *weak constraints* as well, such as the fact that a vertex of high degree is more likely to be in the clique and that if i is not in the clique then its neighbors are less likely and non-neighbors are more likely to be in it.

Indeed, the key conceptual insight of this paper is to phrase the pseudo-calibration property (5.1.2) as a desiderata for our pseudo-distributions. Namely, we say that a function $f = f(G, x)$ is "simple" if it is a low degree polynomial in both the entries of G's adjacency matrix and the variables x, and then require (5.1.2) to hold for all simple functions. It turns out that once you do so, the choice for the pseudo-distribution is essentially determined, and hence proving the main result amounts to showing that it

satisfies the constraints of the SOS algorithm. In the next section we will outline the main ideas of our proof.

Remark 5.1.1 (Planted Clique vs 3XOR). In the light of the discussion above, it is instructive to consider the case of random 3XOR discussed before. Random 3XOR instances on n variables and $\Theta(n)$ constraints are easily seen to be maximally unsatisfiable (that is, at most $\approx 1/2$ the constraints can be satisfied by any assignment) with high probability. On the other hand, Grigorev [59] constructed a sum of squares pseudoexpectation that pretends that such instances instances are satisfiable with high probability, proving a sum of squares lower bound for refuting random 3XOR formulas.

Analogous to the planted distribution $G(n, 1/2, \omega)$, one can define a natural planted distribution over 3XOR instances - roughly speaking, this corresponds to first choosing a random Boolean assignment x^* to n variables and then sampling random 3XOR constraints conditioned on being consistent with x^*. It is not hard to show that pseudo-calibrating with respect to this planted distribution a la (5.1.2) produces precisely the pseudoexpectation that Grigoriev constructed. However, unlike in the planted clique case, in the case of 3XOR, the pseudo-calibration condition implies that for every low-degree monomial x_S, either the value of x_S is completely fixed (if it can be derived via low width resolution from the 3XOR equations of the instance) or it is completely unconstrained.

The pseudoexpectations considered in previous works [47, 85, 43]) are similar to Grigoriev's construction, in the sense that they essentially respect only strong

constraints (e.g., that if A is not a clique in the graph, then the probability that it is contained in the planted clique is zero), but other than that assume that variables are independent. However, unlike the 3XOR case, in the planted clique problem respecting these strong constraints is not enough to achieve the pseudo-calibration condition (5.1.2) and the pseudoexpectation of [47, 85, 43] can be shown to violate weak probabilistic constraints imposed by (5.1.2) even at degree four. See Observation 5.1.4 for an example.

5.1.2 From Calibrated Pseudo-distributions to Sum-of-Squares Lower Bounds

What does Bayesian inference and calibration have to do with Sum-of-Squares? In this section, we show how calibration is almost forced on any pseudodistribution feasible for the Sum-of-Squares algorithm. In order to show that the degree d SOS algorithm fails to certify that a random graph does not contain a clique of size ω, what we need is to show that for a random G, with high probability we can come up with an operator that maps a degree at most d, n-variate polynomial p to a real number $\tilde{\mathbb{E}}_G\, p$ satisfying the following constraints:

1. (Linearity) The map $p \mapsto \tilde{\mathbb{E}}_G\, p$ is linear.
2. (Normalization) $\tilde{\mathbb{E}}_G\, 1 = 1$.
3. (Booleanity constraint) $\tilde{\mathbb{E}}_G\, x_i^2 p = \tilde{\mathbb{E}}\, x_i p$ for every p of degree at most $d-2$ and $i \in [n]$.
4. (Clique constraint) $\tilde{\mathbb{E}}_G\, x_i x_j p = 0$ for every (i, j) that is not an edge and p of degree at most $d-2$.

5. (Size constraint) $\tilde{\mathbb{E}}_G \sum_{i=1}^n x_i = \omega$.

6. (Positivity) $\tilde{\mathbb{E}}_G p^2 \geqslant 0$ for every p of degree at most $d/2$.

Definition 5.1.2. A map $p \mapsto \tilde{\mathbb{E}}_G p$ satisfying the above constraints 1–6 is called a *degree d pseudo-distribution* (w.r.t. the planted clique problem with parameter ω).

We can now restate our main result as follows:

Theorem 5.1.3 (Theorem 3.0.2, restated)**.** *There is some constant c such that if $\omega \leqslant n^{1/2-c(d/\log n)^{1/2}}$ then with high probability over G sampled from $G(n, 1/2)$, there is a degree d pseudodistribution $\tilde{\mathbb{E}}_G$ satisfying constraints 1–6 above.*

Note that all of these constraints would be satisfied if $\tilde{\mathbb{E}}_G p$ was obtained by taking the expectation of p over a distribution on ω-sized cliques in G. However, with high probability there is no $2.1 \log n$-sized clique in G (and let alone a roughly $\sqrt{n}$-sized one) so we will need a completely different mechanism to obtain such a pseudo-distribution.

Previously, the choice of the pseudo-distribution seemed to require a "creative guess" or an "ansatz". For problems such as random 3SAT this guess was fairly natural and almost "forced", while for planted clique as well as some related problems [83] the choice of the pseudo-distribution seemed to have more freedom, and more than one choice appeared in the literature.

For example, Feige and Krauthgamer [47] (henceforth FK) defined a very natural pseudo-distribution $\tilde{\mathbb{E}}^{FK}$ for a weaker hierarchy. For a graph G on n vertices, and subset $A \subseteq [n]$, $\tilde{\mathbb{E}}^{FK}_G x_A$ is equal to zero if A is not a clique in G and equal to $2^{\binom{|A|}{2}} \left(\frac{\omega}{n}\right)^{|A|}$ if A

is a clique, and extended to degree d polynomials using linearity.[4] [47] showed that that for every d, and $\omega < O(\sqrt{n/2^d})$, this pseudo-distribution satisfies the constraints 1–5 as in Definition 5.1.2 as well as a weaker version of positivity (this amounts to the so called "Lovász-Schrijver+" SDP). Meka, Potechin and Wigderson [85] proved that the same pseudo-distribution satisfies all the constraints 1–6 (and hence is a valid degree d pseudo-distribution) as long as $\omega < \tilde{O}(n^{1/d})$. This bound on ω was later improved to $\tilde{O}(n^{1/3})$ for $d = 4$ by [43] and to $\tilde{O}(n^{(\lfloor d/2 \rfloor + 1)^{-1}})$ for a general d by [66].

Interestingly, the FK pseudo-distribution does *not* satisfy the full positivity constraint for larger values of ω. The issue is that while the FK pseudo-distribution satisfies the "strong" constraints that $\tilde{\mathbb{E}}_G^{FK} x_A = 0$ if A is not a clique, it does not satisfy weaker constraints that are implied by (5.1.2). For example, for every constant ℓ, if vertex i participates in $\sqrt{n}$ more ℓ-cliques than the expected number then one can compute that the conditional probability of i belonging in the clique should be a factor $1 + c\omega/\sqrt{n}$ larger for some constant $c > 0$. However, the FK pseudo-distribution does not make this correction. In particular, for every ℓ, there is a simple polynomial that shows that the FK pseudoexpectation is not calibrated.

Observation 5.1.4. Fix $i \in [n]$ and let ℓ be some constant. If $p_G = (\sum_j G_{i,j} x_j)^\ell$ then **(i)** $\mathbb{E}_{G \sim G(n,1/2)} \tilde{\mathbb{E}}_G^{FK} [p_G^2] \leqslant \omega^\ell$ and **(ii)** $\mathbb{E}_{(G,x) \sim G(n,1/2,\omega)} [p_G(x)^2] \geqslant \frac{\omega^{2\ell+1}}{n}$. In particular, when $\omega \gg n^{\frac{1}{\ell+1}}$, $\mathbb{E}_{G \sim G(n,1/2)} \tilde{\mathbb{E}}_G^{FK} [p_G^2] \ll \mathbb{E}_{(G,x) \sim G(n,1/2,\omega)} p_G(x)$.

Proof sketch. For **(ii)** note that with probability (ω/n) vertex i is in the clique, in which case

[4] The actual pseudo-distribution used by [47] (and the followup works [85, 43]) was slightly different so as to satisfy $\tilde{\mathbb{E}}_G (\sum_{i=1}^m x_i)^\ell = \omega^\ell$ for every $\ell \in \{1, \ldots, d\}$. This property is sometimes described as satisfying the constraint $\{\sum_i x_i = \omega\}$.

$\sum_j G_{i,j} x_j = \omega$, and hence the expectation of p_G^2 is at least $(\omega/n)\omega^{2\ell}$. For **(i)**, we open up the expectation and the definition to get (up to a constant depending on ℓ)

$$\sum_{j_1,\dots,j_{2\ell}} G_{i,j_1} \dots G_{i,j_{2\ell}} (\omega/n)^{2\ell} \mathop{\mathbb{E}}_{G\sim G(n,1/2)} 1_{\{i_1,\dots,i_{2\ell}\} \text{ is clique}}$$

Since this expectation is zero unless every variable $G_{i,j}$ is squared, in which case the number of distinct j's is at most ℓ, we can bound the sum by $n^\ell(\omega/n)^\ell = \omega^\ell$. This completes the proof sketch. □

Observation 5.1.4 captures the failure of calibration for a specific polynomial $p_G(x)$ where the coefficients are low-degree functions of the graph G. The polynomial p_G above can be used to show that degree d $\tilde{\mathbb{E}}^{FK}$ does not satisfy the positivity constraint for $\omega \gg n^{1/(\frac{d}{2}+1)}$. This observation is originally due to Kelner, see [66]

Fact 5.1.5. *Let p_G be as in the Observation 5.1.4. Then, there exists a C such that for $q = q_G = (C\omega^\ell x_S - p_G)$ with high probability over the graph $G \sim G(n, 1/2)$, $\tilde{\mathbb{E}}^{FK}[q_G^2] < 0$ for $\omega \gg n^{\frac{1}{\ell+1}}$.*

For the case $d = 4$, Hopkins et al [65] proposed an "ad hoc" fix for the FK pseudo-distribution that satisfies positivity up to $\omega = \tilde{O}(\sqrt{n})$, by explicitly adding a correction term to essentially calibrate for the low-degree polynomials q_G from Fact 5.1.5. However, their method did not extend even for $d = 6$, because of the sheer number of corrections that would need to be added and analyzed. Specifically, there are multiple families of polynomials such that their $\tilde{\mathbb{E}}^{FK}$ value departs significantly from their calibrated value in expectation and gives multiple points of failure of positivity in a manner similar to Observation 5.1.4 and Fact 5.1.5. Moreover, "fixing" these families by the correction as in

case of degree four leads to new families of polynomials that fail to achieve their calibrated value and exhibit negative pseudoexpectation for their squares and so on.

The *coefficients* of the polynomial p_G of Observation 5.1.4 are themselves low degree polynomials in the adjacency matrix of G. This turns out to be a common feature in all the families of polynomials one encounters in the above works. Thus our approach is to fix all these polynomials *by fiat*, by placing the constraint that the pseudo-distribution must satisfy (5.1.2) for every such polynomial, and using that as our implicit definition of the pseudo-distribution. Indeed it turns our that once we do so, the pseudo-distribution is essentially determined. Moreover, (5.1.2) guarantees that it satisfies many of the "weak global constraints" that can be shown using Bayesian calculations.

Ultimately we will construct the map $G \mapsto \tilde{\mathbb{E}}_G$ as a low degree polynomial in G. Why is it OK to make such a restriction? One justification is the heuristic that the pseudo-distribution itself must be simple since we know that it is efficiently computable (via the SOS algorithm) from the graph G. Another justification is that by forcing the pseudo-distribution to be low-degree we are essentially making it *smooth* or "high entropy", which is consistent with the Jaynes *maximum entropy principle* [72, 71]. Most importantly — and this is the bulk of the technical work of this paper and the subject of the next subsection — this pseudo-distribution can be shown to satisfy *all* the constraints 1–6 of Definition 5.1.2 including the positivity constraint.

We believe that this principled approach to designing pseudo-distributions elucidates the power and limitations of the SOS algorithm in cases such as planted clique, where accounting for weak global correlations is a crucial aspect of the problem.

Remark 5.1.6 (*Where does the planted distribution arise from?*). Theorem 5.1.3 (as well as Theorem 3.0.2) makes no mention of the planted distribution $G(n, 1/2, \omega)$ and only refers to an actual random graph. Thus it might seem strange that we base our pseudo-distribution on the planted distribution via (5.1.2). One way to think about the planted distribution is that it corresponds to a *Bayesian prior* distribution on the clique. Note that this is the *maximum entropy* distribution on cliques of size ω, and so it is a natural choice for a prior per Jaynes's principle of maximum entropy. Our actual pseudo-distribution can be viewed as correcting this planted distribution to a posterior that respects simple inferences from the observed graph G.

5.1.3 Towards Proving Positivity: Structure vs. Randomness

We have seen that pseudo-calibration is desirable both *a priori* and in light of the failure of previous lower-bound attempts. Now we turn to the question: How do we formally define a pseudo-calibrated linear map $\tilde{\mathbb{E}}_G$, and show that it satisfies constraints 1–6 with high probability, to yield Theorem 5.1.3?

We will require (5.1.2) to hold with respect to every function $f = f(G, x)$ that has degree at most τ in the entries of the adjacency matrix G and degree at most d in the variables x, and in addition we require that the map $G \mapsto \tilde{\mathbb{E}}_G$ is itself of degree at most τ in G, then this completely determines $\tilde{\mathbb{E}}_G$. For any $S \subseteq [n]$, $|S| \leqslant d$, using the Fourier

transform we can write $\tilde{\mathbb{E}}_G[x_S]$ as an explicit low degree polynomial in G_e:

$$\tilde{\mathop{\mathbb{E}}_G}[x_S] = \sum_{\substack{T \subseteq \binom{[n]}{2} \\ |\mathcal{V}(T) \cup S| \leqslant \tau}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(T)\cup S|} \chi_T(G), \tag{5.1.3}$$

where $\mathcal{V}(T)$ is the set of nodes incident to the subset of edges (i.e., graph) T and $\chi_T(G) = \prod_{e \in T} G_e$. We carry out this computation in Section 5.4. For $\omega \approx n^{0.5-\varepsilon}$, we will need to choose the truncation threshold $\tau \gtrapprox d/\varepsilon$. It turns out that constraints 1–5 are easy to verify and thus we are left with proving the *positivity constraint*. Indeed this is not surprising as verifying this constraint is always the hardest part of a Sum-of-Squares lower bound.

As is standard, to analyze this positivity requirement we work with the *moment matrix* of $\tilde{\mathbb{E}}_G$. Namely, let $\mathcal{M}$ be the $\binom{n}{\leqslant d/2} \times \binom{n}{\leqslant d/2}$ matrix where $\mathcal{M}(I, J) = \tilde{\mathbb{E}}_G \prod_{i \in I} x_i \prod_{j \in J} x_j$ for every pair of subsets $I, J \subseteq [n]$ of size at most $d/2$. Our goal can be rephrased as showing that $\mathcal{M} \succeq 0$ (i.e., $\mathcal{M}$ is positive semidefinite).

Given a (symmetric) matrix N, to show that $N \succeq 0$ our first hope might be to diagonalize N. That is, we would hope to find a matrix V and a diagonal matrix D so that $N = VDV^\top$. Then as long as every entry of D is nonnegative, we would obtain $N \succeq 0$. Unfortunately, carrying this out directly can be far too complicated. Even the eigenvectors of simple random matrices are not completely understood, let alone matrices like ours with intricate dependencies among the entries. However, as the next example demonstrates, it is sometimes possible to prove positivity for a random matrix using what we call *approximate* diagonalization.

Example: Planted Clique Lower Bound for $d = 2$ (a.k.a. Basic SDP). Consider the problem of producing a pseudo-distribution $\tilde{\mathbb{E}}$ satisfying constraints 1–6 of Definition 5.1.2,

with $d = 2$. In this simple case, many subtleties can be safely be ignored, but can still provide some intuition. For $d = 2$, it is enough to define $\tilde{\mathbb{E}}\, x_i$ and $\tilde{\mathbb{E}}\, x_i x_j$ for every $i \in [n]$ and $\{i, j\} \subseteq [n]$. Let $\tilde{\mathbb{E}}\, x_i = (\omega/n)$ for every i, and set $\tilde{\mathbb{E}}\, x_i x_j$ to be $\left(\frac{\omega}{n}\right)^2$ if (i, j) is an edge in G and zero otherwise. It is not hard to show that positivity reduces to showing that $\mathcal{N} \succeq 0$ where $\mathcal{N}$ is the $n \times n$ matrix with $\mathcal{N}_{i,j} = \tilde{\mathbb{E}}\, x_i x_j$. Using standard results on random matrices, $\mathcal{N}$ has one eigenvalue (whose corresponding eigenvector is close to the vector $u = (1/\sqrt{n}, \ldots, 1/\sqrt{n})$) of value ω^2/n, while all others are distributed in the interval $\frac{\omega}{n} \pm O\left(\frac{\omega^2}{n^2}\sqrt{n}\right)$ which is strictly positive as long as $\omega \ll \sqrt{n}$. Thus, while we cannot explicitly diagonalize $\mathcal{N}$, we have enough information to conclude that it is positive semidefinite. In other words, it was enough for us to get an *approximate diagonalization* for $\mathcal{N}$ of the form $\mathcal{N} \approx \frac{\omega^2}{n} u u^\top + \frac{\omega}{n} Id + E$ for some sufficiently small (in spectral norm) "error matrix" E. Ultimately we will need to do something similar, but with many eigenvalues and many error matrices that are inter-dependent.

Approximate Factorization for $\mathcal{M}$. We return now to the moment matrix $\mathcal{M}$ for our (pseudo)calibrated pseudodistribution. Our goal is to give an approximate diagonalization of $\mathcal{M}$. There are several obstacles to doing so:

1. In the case $d = 2$ there was just one rank-1 approximate eigenspace to be handled. The number of these approximate eigenspaces will grow with d, so we will need a more generic way to handle them.

2. Each approximate eigenspace corresponds to a family of polynomials $\{p\}$ whose calibrated pseudoexpectations are all roughly equal. (In the case $d = 2$, the only

interesting polynomial was the polynomial $\sum_j x_j$ whose coefficients are proportional to the vector $u = (1/\sqrt{n}, \ldots, 1/\sqrt{n})$.) As we saw in Observation 5.1.4, if p_G is a polynomial whose coefficients depend on the graph G, even in simple ways, the calibrated value $\tilde{\mathbb{E}}_G\, p_G$ may also depend substantially on the graph. Thus, when we write $\mathcal{M} \approx \mathcal{L}\mathcal{Q}\mathcal{L}^\top$ for some approximately-diagonal matrix $\mathcal{Q}$, we will need the structured part $\mathcal{L} = \mathcal{L}(G)$ to itself be graph-dependent.

3. The errors in our diagonalization of $\mathcal{M}$ — corresponding in our $d = 2$ example to the matrix E — will not be small enough to ignore as we did above. Instead, each error matrix will itself have to be approximately diagonalized, recursively until these errors are driven down sufficiently far in magnitude.

We now discuss at a high level our strategy to address items (1) and (2). The resolution to item (3) is the most technical element of our proof, and we leave it for later. Consider the vector space of all polynomials $f : \{0,1\}^{\binom{n}{2}} \times \mathbb{R}^n \to \mathbb{R}$ which take a graph and an n-dimensional real vector and yield a real number. (We write $f_G(x)$, where G is the graph and $x \in \mathbb{R}^n$.) If we restrict attention to the subspace of those of degree at most d in x, we obtain the polynomials in the domain of our operator $\tilde{\mathbb{E}}_G$. If we additionally restrict to the subspace of polynomials which are low degree in G, we obtain the family of polynomials so that $\mathbb{E}_G\, \tilde{\mathbb{E}}_G\, f_G(x)$ is calibrated. Call this subspace $\mathcal{V}$.

Our goal would to be find an approximate diagonalization for all the non-trivial eigenvalues of $\mathcal{M}$ using only elements from $\mathcal{V}$. The advantage of doing so is that for every $f \in \mathcal{V}$, we can calculate $\mathbb{E}_G\, \tilde{\mathbb{E}}_G\, f_G^2$ using the pseudo-calibration condition (5.1.2). In particular it means that if we find a function f such that f_G is with high probability an ap-

proximate eigenvector of G, then we can compute the corresponding expected eigenvalue $\lambda(f)$.

A crucial tool in finding such an approximate eigenbasis is the notion of *symmetry*. For every f, if f' is obtained from f via a permutation of the variables $x_1, \ldots, x_n$, then $\mathbb{E}_G \tilde{\mathbb{E}}_G f_G^2 = \mathbb{E}_G \tilde{\mathbb{E}}_G f_G'^2$. The result of this symmetry, for us, is that our approximate diagonalization requires only a constant (depending on d) number of eigenspaces. This argument allows us to restrict our attention to a constant number of classes of polynomials, where each class is determined by some finite graph U that we call its *shape*. For every polynomial f with shape U, we compute (approximately) the value of $\mathbb{E}_G \tilde{\mathbb{E}}_G f_G^2$ as a function of a simple combinatorial property of U, and our approximate eigenspaces correspond to polynomials with different shapes.

We can show that that in expectation our approximate eigenspaces will have non-negative eigenvalues since the pseudo-calibration condition (5.1.2) in particular implies that for every f that is low degree in both G and x, $\mathbb{E}_G \tilde{\mathbb{E}}_G f_G^2 \geqslant 0$. However, the key issue is to deal with the error terms that arise from the fact that these are only approximate eigenspaces. One could hope that, like in other "structure vs. randomness" partitions, this error term is small enough to ignore. Alas, this is not the case, and we need to handle it recursively, which is the crux of item (3) and the cause of much of the technical complications of our paper.

Remark 5.1.7 (*Structure vs. randomness*). At a high level our approach can be viewed as falling into the general paradigm of "structure vs. randomness" as discussed by Tao [102]. The general idea of this paradigm is to separate an object O into

a "structured" part that is simple and predictable, and a "random" part that is unpredictable but has small magnitude or has some global statistical properties.

One example of this is the Szemerédi regularity lemma [101] as well variants such as [51] that partition a matrix into a sum of a low rank and pseudorandom components. Another example arises from the random models for the *primes* (e.g., see [4, 57]). These can be thought of positing that, as far as certain simple statistics are concerned, (large enough) primes can be thought of as being selected randomly conditioned on not being divisible by $2, 3, 5$ etc.. up to some bound w.

All these examples can be viewed from a computationally bounded Bayesian perspective. For every object O we can consider the part of O that can be inferred by a computationally bounded observer to be O's *structured* component, while the remaining uncertainty can be treated as if it is *random*, even if in actuality it is fully determined. Thus in our case, even though for almost every particular graph G from $G(n, 1/2, \omega)$, the clique x is fully determined by G, we still think of x as having a "structured" part which consists of all the inferences a "simple" observer can make from G (e.g., that if i and j are non-neighbors then $x_i x_j = 0$), and a "random" part that consists of the remaining uncertainty. As in other cases of applying this paradigm, part of the technical work is bounding the magnitude (in our case in spectral norm) that arises from the "random" part, though as mentioned above in our case we need a particularly delicate control of the error terms which ends up causing much of the technical difficulty.

5.2 Proving Positivity: A Technical Overview

We now discuss in more detail how we prove that the *moment matrix* $\mathcal{M}$ corresponding to our pseudo-distribution is positive semidefinite. Recall that this is the $\binom{n}{\leqslant d/2} \times \binom{n}{\leqslant d/2}$ matrix $\mathcal{M}$ such that $\mathcal{M}(I,J) = \tilde{\mathbb{E}}_G \prod_{i\in I} x_i \prod_{j\in J} x_j$ for every pair of subsets $I, J \subseteq [n]$ of size at most $d/2$, and that it is defined via (5.1.3) as

$$\mathcal{M}(I,J) = \sum_{\substack{T\subseteq\binom{[n]}{2} \\ |\mathcal{V}(T)\cup I\cup J|\leqslant\tau}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(T)\cup I\cup J|} \chi_T(G)\ . \tag{5.2.1}$$

The matrix $\mathcal{M}$ is generated from the random graph G, but its entries are *not* independent. Rather, each entry is a polynomial in G_e, and there are some fairly complex dependencies between different them. Indeed, these dependencies will create a spectral structure for $\mathcal{M}$ that is very different from the spectrum of standard random matrices with independent entries and makes proving that $\mathcal{M}$ is positive semidefinite challenging. Our approach to showing that $\mathcal{M}$ is positive semidefinite is through a type of "symbolic factorization" or "approximate diagonalization," which we explain next.

5.2.1 Warm Up

It is instructive to begin with the tight analysis presented in [66] of the moments constructed in [85][5]. These moments can in fact obtained by using truncation threshold $\tau = |S|$ in (5.1.3). This choice of τ is the smallest possible for which the resulting construction satisfies the hard clique constraints. [66] show that this construction satisfies positivity

[5]The construction in [85] actually also satisfies $\sum x_i = \omega$ as a constraint which causes the precise form to differ. We ignore this distinction here.

for $\omega \lessapprox n^{1/(\frac{d}{2}+1)}$.

For the purpose of this overview, let us work with the principal submatrix F indexed by subsets I and J of size exactly d. The analysis in [66] proceeds by first splitting F into $d+1$ components $F = F_0 + F_1 + \cdots + F_d$ where $F_i(I, J) = F(I, J)$ if $|I \cap J| = i$ and 0 otherwise. Below, we discuss two of the key ideas involved that will serve as an inspiration for us.

As discussed before, we must approximately diagonalize the matrix F in the sense that the off diagonals blocks must be "small enough" to be charged to the on diagonal block. Thus the main question before us is obtain an (approximate) understanding of the spectrum of F that allows us to come up with a "change of basis" in which the off diagonal blocks are small enough to be charged to the positive eigenmass in the on-diagonal blocks.

Let us consider the piece F_0 for our discussion here. As alluded to in Section 5.2, we want to break F into minimal pieces so that each piece is symmetric under the permutation of vertices. We can hope that each piece will then essentially have a single dominating eigenvalue that can be determined relatively easily. Below, we will essentially implement this plan.

First, we need to decide what kind of "pieces" we will need. These are the *graphical matrices* that we define next.

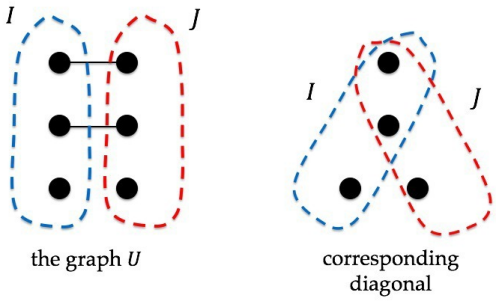
Definition 5.2.1 (Graphical Matrices (see Def 5.6.6 for a formal version))**.** Let U be a graph on $[2d]$ with specially identified subsets left and right subsets $[d]$ and $[2d] \setminus [d]$. For any $I, J \in \binom{[n]}{d}$, $I \cap J = \emptyset$, let $\pi_{I,J}$ be an injective map that takes $[d]$ into I and $[2d] \setminus [d]$ into J using a fixed convention. The graphical matrix M_U with graph U is then defined by $M_U(I, J) = \chi_{\pi_{I,J}(U)}(G)$.

The starting point of the analysis is to decompose $F_0 = \sum_U \left(\frac{\omega}{n}\right)^{2d} M_U$, where M_U is the graphical matrix with shape U. Graphical matrices as above turn out to be the right building blocks for spectral analysis of our moment matrix. This is because a key observation in [66] shows that a simple combinatorial parameter, the size of the maximum bipartite matching between the left and right index in U (i.e. between $[d]$ and $[2d] \setminus [d]$), determines the spectral norm of M_U. Specifically, when U has a maximum matching of size $t < d$, the spectral norm of M_U is $\tilde{O}(n^{d-\frac{t}{2}})$, with high probability. Observe that when $d = 2$ and U is a single edge connecting the left vertex with the right, M_U is just the $\{-1, 1\}$-adjacency matrix of the underlying random graph and it is well known that the spectral norm in this case is $\Theta(\sqrt{n})$ matching the more general claim above.

In particular, this implies that when U has a perfect matching, M_U is pseudorandom in the sense that F_U essentially has the spectral norm $\approx n^{d/2}$, the same as that of an independent $\{-1, 1\}$ random matrix of the same dimensions. This allows M_U to be bounded against the positive eigenvalue $\left(\frac{\omega}{n}\right)^d$ of the diagonal matrix F_d as $\left(\frac{\omega}{n}\right)^d \gg \left(\frac{\omega}{n}\right)^{2d} n^{d/2}$ (even for ω approaching $\sqrt{n}$!). However for M_U when U has a maximum matching of size $t < d$, one can't bound against the diagonal matrix F_d anymore.

The next main idea is to note that for every M_U there's an appropriate "diagonal" against which we must charge the negative eigenvalues of M_U. When U has a perfect matching, this is literally the diagonal matrix F_d as done above. However, when, say, U is a (bipartite) matching of size $t < d$, we should instead charge against the "diagonal" matrix that can thought of as obtained by "collapsing" each matching edge into a vertex in U. In particular, this collapsing produces a matrix that lies in the decomposition of F_t.

There are a two main takeaways from this analysis that would serve as inspiration in



the analysis of our actual construction. First is the decomposition into graphical matrices in order to have a coarse handle on the spectrum of the moment matrix. Second, the "charging" of negative eigenvalues against appropriate "diagonals" is essentially governed by the combinatorics of matchings in U.

5.2.2 The Main Analysis

We can now try to use the lessons from the warm up analysis to inspire our actual analysis. To begin with, we recall that each graphical matrix was obtained by choosing an appropriate (set of) Fourier monomials for any entry indexed by I, J. However, since for our actual construction we have monomials of much higher degree, we need to extend the notion of graphical matrices with *shapes* corresponding to larger graphs U. See Def 5.6.6 for a formal definition.

It turns out that the right combinatorial idea to generalize the size of the maximum matching and control the spectral norm of the graphical matrices M_U is the maximum number of *vertex disjoint paths* between specially designated left and right endpoints of U (themselves the generalization of the bipartition we had in the warmup). Using Menger's

theorem, this is equal to the size of a minimal collection of vertices that separates the left and right sets in the graph U, which we call the *separator size* of U.

Finally, we need a "charging" argument to work with the approximate diagonalization we end up with. Generalizing the idea in the warm up here is the hardest part of our proof, but relates again to the notion of vertex separators defined above. In the warm up, we used a naive charging scheme, breaking the moment matrix into simpler (graphical) matrices, each of which was either a "positive diagonal" mass or a "negative off-diagonal mass", and pairing up the terms. Such a crude association doesn't work out immediately in the general setting. Instead, large groups of graphical matrices must be treated all at once. In each subspace of our approximate diagonalization of the moment matrix $\mathcal{M}$, we collect the "positive diagonal mass" and the "negative off digonal mass" that needs to be charged to it together and build an approximately PSD matrix out of it. As alluded to before, the error in this approximation is not negligible and thus we must further recurse on the error terms. In what follows, we discuss the factorization process that accomplishes the charging scheme implicitly and the recursive factorization for the error terms in some more detail. Consider some graph $T \subseteq \binom{[n]}{2}$, that corresponds to one term in the sum in (5.2.1) above, and let q be the minimum size of a set that separates I from J in T. Such a set is not necessarily unique but we can define the *leftmost* separator $\mathsf{left-sep}(T) = S_\ell$ to be the q-sized separator that is closest to I and the *rightmost* separator $\mathsf{right-sep}(T) = S_r$ to be the q-sized separator that is closest to J.

We can rewrite the (I, J) entry moment matrix $\mathcal{M}$ (5.2.1) by collecting monomials T with a fixed choice of the leftmost and rightmost separators S_ℓ and S_r. This step corresponds to collecting terms with similar spectral norms together accomplishing the

goal of collecting together into a term, the "positive diagonal mass" and the "negative off diagonal mass" that are implicitly charged to each other in the intended approximate diagonalization.

$$\mathcal{M}(I,J) = \sum_{1 \leqslant q \leqslant |I|,|J|} \sum_{S_\ell, S_R : |S_\ell| = |S_r| = q} \sum_{\substack{T \subseteq \binom{[n]}{2} \\ |\mathcal{V}(T) \cup I \cup J| \leqslant \tau \\ \text{left-sep}(T) = S_\ell, \text{right-sep}(T) = S_r}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(T) \cup I \cup J|} \chi_T(G) \tag{5.2.2}$$

We can then partition T into three subsets $\mathcal{R}_\ell$, $\mathcal{R}_m$ and $\mathcal{R}_r$ that represent the part of the graph T between I and S_ℓ, the part between S_ℓ and S_r and the part between S_r and J respectively (where edges within S_ℓ and edges within S_r are all placed in $\mathcal{R}_m$, see Definition 5.5.4). We thus immediately obtain that

$$\chi_T(G) = \chi_{\mathcal{R}_\ell}(G) \chi_{\mathcal{R}_m}(G) \chi_{\mathcal{R}_r}(G) \ .$$

Thus:

$$\mathcal{M}(I,J) = \sum_{1 \leqslant q \leqslant |I|,|J|} \sum_{S_\ell, S_R : |S_\ell| = |S_r| = q} \sum_{\substack{T \subseteq \binom{[n]}{2} \\ |\mathcal{V}(T) \cup I \cup J| \leqslant \tau \\ \text{left-sep}(T) = S_\ell \\ \text{right-sep}(T) = S_r}} \left(\left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_\ell)|} \chi_{\mathcal{R}_\ell}(G)\right) \left(\left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_m)| - 2q} \chi_{\mathcal{R}_m}(G)\right) \left(\left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_r)|} \chi_{\mathcal{R}_r}(G)\right) \tag{5.2.3}$$

One could hope that we could replace the RHS of (5.2.3) by

$$\sum_{\substack{1\leqslant q\leqslant |I|,|J| \\ \tau_1+\tau_2+\tau_3\leqslant\tau}} \sum_{\substack{S_\ell\subseteq\binom{[n]}{q} \\ S_r\subseteq\binom{[n]}{q}}} \left(\sum_{\substack{\mathcal{R}_\ell \\ \mathcal{V}(\mathcal{R}_\ell)\supseteq I\cup S_\ell \\ |\mathcal{V}(\mathcal{R}_\ell)|=\tau_1}} \left(\tfrac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_\ell)|} \chi_{\mathcal{R}_\ell}(G) \right) \left(\sum_{\substack{\mathcal{R}_m \\ \mathcal{V}(\mathcal{R}_m)\supseteq S_\ell\cup S_r \\ |\mathcal{V}(\mathcal{R}_m)|=\tau_2}} \left(\tfrac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_m)|-2q} \chi_{\mathcal{R}_m}(G) \right) \left(\sum_{\substack{\mathcal{R}_r \\ \mathcal{V}(\mathcal{R}_r)\supseteq S_r\cup J \\ |\mathcal{V}(\mathcal{R}_r)|=\tau_3}} \left(\tfrac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_r)|} \chi_{\mathcal{R}_r}(G) \right) \tag{5.2.4}$$

In fact, it turns out we can focus attention (up to sufficiently small error in the spectral norm) to the case $\tau_1 \leqslant \tau/3$, $\tau_2 \leqslant \tau/3$, $\tau_3 \leqslant \tau/3$ in which case if $M(I,J)$ was equal to (5.2.4) we could simply write

$$\mathcal{M} = \sum_q \mathcal{L}_q \mathcal{Q}_q \mathcal{L}_q^\dagger$$

where for $I, S \subseteq [n]$ with $|I| \leqslant d$ and $|S| = q$, we let $\mathcal{L}_q(I,S)$ be the sum of $(\omega/n)^{|V(\mathcal{R}_\ell)|}\chi_{\mathcal{R}_\ell}(G)$ over all graphs $\mathcal{R}_\ell$ of at most $\tau/3$ vertices connecting I to S, and for S, S' of size q, we let $\mathcal{Q}_q(S,S')$ be the sum of $(\omega/n)^{|\mathcal{R}_m|-2q}\chi_{\mathcal{R}_m}(G)$ over all graphs $\mathcal{R}_m$ of at most $\tau/3$ vertices connecting S to S'.

Thus, in this case, this reduces our task of showing that $\mathcal{M}$ is positive semidefinite to showing that for every q, the matrix $\mathcal{Q} = \mathcal{Q}_q$ is positive semidefinite. However the main complication is that there are cross terms in the product $\mathcal{L}_q\mathcal{Q}_q\mathcal{L}_q^\top$ that correspond to repeating the same vertex (not in S_ℓ and S_r) in more than one of $\mathcal{R}_\ell, \mathcal{R}_m$ and $\mathcal{R}_r$. There is no matching term in the Fourier decomposition of $\mathcal{M}(I,J)$. So at best, for every fixed q, we can write the part of $\mathcal{M}$ corresponding to indices I, J with minimal vertex separator

equal to q as

$$\mathcal{L}Q_0\mathcal{L}^\top - \mathcal{E}_1$$

for some error matrix $\mathcal{E}_1$ that exactly cancels out the extra terms contributed by cross terms with repeated vertices. Unfortunately, the spectral norm of this error matrix $\mathcal{E}_1$ is not small enough that we could simply ignore it. Luckily however, we can recurse and factorize $\mathcal{E}_1$ approximately as well. We can form a new graph T' by taking the parity of the edge sets in $\mathcal{R}_\ell$, $\mathcal{R}_m$ and $\mathcal{R}_r$. Now we find the leftmost and rightmost separators that separate I and J from each other, and from all repeated vertices. This gives us another decomposition of a graph into three pieces, from which we can write

$$\mathcal{E}_1 = \mathcal{L}Q_1\mathcal{L}^\top - \mathcal{E}_2$$

for some other matrix Q_1. Continuing this argument gives us for every q a factorization of $\mathcal{M}_q$ as

$$\mathcal{L}(Q_0 - Q_1 + Q_2 - \ldots - Q_{2d-1} + Q_{2d})\mathcal{L}^\top - (\xi_0 - \xi_1 + \xi_2 - \ldots - \xi_{2d-1} + \xi_{2d})$$

The error matrices $\xi_0, \xi_1, \ldots, \xi_{2d}$ arise from truncation issues, which we have ignored in the argument above and turn out to be negligible.

It is not hard to show that $Q_0 \succeq D$ for some positive semidefinite matrix D that we define later. What remains is to bound the remaining matrices $Q_1, \ldots Q_{2d-1}$ in order to conclude that $\mathcal{M}$ is positive semidefinite. Next, we elaborate on the structure of these matrices. It turns out that we can define the "shape" of a graph $\mathcal{R}_m$ in an appropriate way so that

$$Q_i^U(S_\ell, S_r) = \sum_{\text{shape}(\mathcal{R}_m)=U} c_i(\mathcal{R}_m)\chi_{\mathcal{R}_m}$$

where U is a finite (for constant d) sized graph with vertex set $A \cup B \cup C$, where we call A the "left" side of U and B the "right" side of U. Moreover $Q_i = \sum_U Q_i^U$. Now Q_i^U is a random matrix and special cases of this general family of matrices (for particular choices of U) arise in several earlier works on lower bounds for planted clique. Medarametla and Potechin [84] showed that the spectral norm of Q^U can be controlled by a bound on its coefficients and a few combinatorial parameters of U — namely $|\mathcal{V}(U)|$, $|A \cap B|$ and the number of vertex disjoint paths between A/B and B/A.

A major challenge in our work is to understand and analyze the coefficients c_i. In the course of decomposing $\mathcal{M}$, we are able to characterize $c_i(\mathcal{R}_m)$ as an appropriately weighted sum over $c_{i-1}(\mathcal{R}'_m)$ where $\mathcal{R}'_m$ ranges over the middle piece of all graphs with leftmost and rightmost separators S_ℓ and S_r that could have resulted in $\mathcal{R}_m$ due to repeated vertices. Recall that when there are repeated vertices, we take the parity of the edge sets of the three pieces and compute a new set of left and rightmost vertex separators. The set of $\mathcal{R}'_m$'s that could result in $\mathcal{R}_m$ is complicated. Instead, our approach is to show that the various combinatorial parameters of $\mathcal{R}'_m$ (which affect the spectral norm bounds) tradeoff against each other when accounting for the effect of repeated vertices. This allows us to bound their contribution and ultimately show that the coefficients c_i decay quickly enough for all values of $\omega < n^{1/2-\varepsilon}$ that we can bound each Q_i for $i > 1$ as $-\frac{D}{8d} \succeq Q_i \succeq \frac{D}{8d}$, and this completes our proof.

5.3 Preliminaries

5.3.1 General Notation

- We use small Greek letters indicate constants/parameters.

- $\mathbb{P}_d^n$ denotes the linear space of all *multilinear* polynomials of degree at most d on $\{0,1\}^n$.
- We write $\mathbb{1}_Q$ for any event Q to be the 0-1 indicator of whether Q happens.
- For a subset $T \subseteq \binom{[n]}{2}$ of edges of a graph on vertex set $[n]$, we write $\mathcal{V}(T) \subseteq [n]$ to denote the vertices that have at least one edge incident on them in T.
- For a matrix $Q \in \mathbb{R}^{N \times N}$, $\|Q\|$ denotes its spectral norm (or the largest singular value) and $\|Q\|_F = \sqrt{\sum_{x,y \in [N]} Q(x,y)^2}$ denotes its Frobenius norm.
- For a graph G, let $C_q = C_q(G) = \{I \subseteq [n] : I \text{ is a } q\text{-clique in } G\}$, and let $C_{\leqslant q} = \bigcup_{q' \leqslant q} C_{d'}$. Let $C(G) = C_{\leqslant \infty}$ be the collection of all cliques in G. We count the empty set and all singletons as cliques.
- We write $\mathcal{G}(n, \frac{1}{2})$ to denote the distribution on graphs on the vertex set $[n]$ where each edge is included with probability $1/2$ independently of others.
- We say that an event E with respect to the probability distribution $\mathcal{G}(n, \frac{1}{2})$ happens *with high probability (w.h.p.)* if $\mathbb{P}[E] \geqslant 1 - \Omega(1)/n^{10 \log n}$ for large enough n.
- We write $f(n) \ll g(n)$ to mean that for every constant c there is an n_0 such that if $n \geqslant n_0$, $f(n) \leqslant Cg(n)$.

5.3.2 Graphs

We identify a graph G with its $\{-1,1\}$ adjacency matrix and write $G_e \in \{-1,1\}$ for the $\{-1,1\}$-indicator of whether $e \in [n] \times [n]$ is an edge (indicated by $G_e = +1$) in the graph G or not. When $G \sim \mathcal{G}(n, \frac{1}{2})$, G_e are independent $\{-1,1\}$-random variables.

A *graph function* is a real-valued function of the variables $G_e \in \{-1, 1\}$ for $e \in \binom{[n]}{2}$. For graphs $G^1, G^2, \ldots, G^k$ on the vertex set $[n]$, we define $\Delta(G^1, G^2, \ldots, G^k)$ to be the graph G satisfying $G_e = \Pi_{i \leqslant k} G^i_e$.

Definition 5.3.1 (Vertex Separator)**.** For a graph G on $[n]$ and vertex sets $I, J \subseteq [n]$, a set of vertices $S \subseteq [n]$ is said to be a *minimal vertex separator* if S is a set of smallest possible size such that every path between I and J in G passes through some vertex of S.

Often, I and J will be allowed to intersect in which case any vertex separator must contain $I \cap J$.

Fact 5.3.2 (Menger's Theorem)**.** *For a graph G on $[n]$ and two subsets of vertices $I, J \subseteq [n]$, the maximum number of vertex disjoint paths between I and J in G is equal to the size of any minimal vertex separator between I and J in G.*

5.3.3 Fourier Analysis

Any graph function $f : G \to \mathbb{R}$ can be represented as a Fourier polynomial in the variables G_e:

$$f(G) = \sum_{W \subseteq \binom{[n]}{2}} \hat{f}(W) \chi_W(G),$$

where $\chi_W(G)$ is the *parity* function on edges in W:

$$\chi_W(G) = \Pi_{e \in W} G_e.$$

The parity function χ_W are an orthonormal basis for functions on G under the inner product defined by $\langle f, h \rangle = \mathbb{E}_{G \sim G(n, \frac{1}{2})}[f(G)h(G)]$ for any graph functions f and h.

The following fact is easy to verify:

Fact 5.3.3. *Let G be a graph on n described by the vector $G \in \{-1,1\}^{\binom{n}{2}}$. For any subset $S \subseteq [n]$ of the vertices, we have the identity:*

$$\sum_{W \subseteq \binom{S}{2}} \chi_W(G) = \begin{cases} 2^{\binom{|S|}{2}} & \text{if } S \text{ is a clique in } G, \\ 0 & \text{otherwise.} \end{cases}$$

5.3.4 The Sum-of-Squares Algorithm

The sum of squares algorithm has several equivalent definitions. We follow the notation of *pseudoexpectations* as in the survey of Barak and Steurer [28].

Definition 5.3.4 (Pseudoexpectation)**.** A linear operator $\tilde{\mathbb{E}} : \mathbb{P}_d^n \to \mathbb{R}$ is said to be a *degree d-pseudoexpectation* if it satisfies:

1. Normalization: $\tilde{\mathbb{E}}[\mathbf{1}] = 1$.
2. Positive Semidefiniteness: $\tilde{\mathbb{E}}[p^2] \geqslant 0$ for every polynomial $p \in \mathbb{P}_d^n$.

A pseudoexpectation operator $\tilde{\mathbb{E}}$ on $\mathbb{P}_d^n$ is said to satisfy a constraint $\{p = 0\}$ for any $p \in \mathbb{P}_d^n$ if for every polynomial $q \in \mathbb{P}_d^n$ such that $p \cdot q \in \mathbb{P}_d^n$, $\tilde{\mathbb{E}}[pq] = 0$.

Given a set of constraints $\{p_i = 0\}$ for $1 \leqslant i \leqslant m$ and an objective polynomial p, degre sum of squares algorithm of degree d solves the problem

$$\arg\max \tilde{\mathbb{E}}[p]$$

over all degree d pseudoexpectations $\tilde{\mathbb{E}}$ that satisfy $\{p_i = 0\}$ for $1 \leqslant i \leqslant m$.

5.4 The Pseudo-expectation

We now define our pseudo-distribution operator $\tilde{\mathbb{E}}_G$. As discussed in Section 5.1.2, it is based on requiring (5.1.2) to hold for every f that has degree at most τ in G and d in x.

Important Parameters. The following parameters will be fixed for the rest of the paper.

- $\varepsilon \in (0, 1/2)$, which determines the size $\omega = n^{1/2-\varepsilon}$ of the planted clique.
- $d = d(n) \in \mathbb{N}$, the degree of the SoS relaxation against which we prove a lower bound.
- $\tau = \tau(n) \in \mathbb{N}$, the degree of our pseudoexpectation $\tilde{\mathbb{E}}$ as a function of $G \sim G(n, 1/2)$.

We always assume that $Cd/\varepsilon \leqslant \tau \leqslant (\varepsilon/C)\log n$ and $\varepsilon \geqslant C \log\log n / \log n$ for a sufficiently-large constant C. Eventually we will set $d = (\varepsilon/C)^2 \log n$, (this yields the parameters stated in Theorem 3.0.2, since then $n^{1/2-\varepsilon} = n^{1/2-\Omega(d/\log n)^{1/2}}$), which implies that $\varepsilon \gg \log\log n / \log n$.

5.4.1 Definition of $\tilde{\mathbb{E}}$

As discussed previously, $\tilde{\mathbb{E}}$ is completely specified by its *multilinear moments*: $\tilde{\mathbb{E}}[x_I]$ for $I \subseteq [n]$ and $|I| \leqslant d$. $\tilde{\mathbb{E}}[x_I]$ is a function of G_e for $e \in \binom{[n]}{2}$ and can be written as a polynomial in G_e with coefficients $\widehat{\tilde{\mathbb{E}}[x_I]}(T)$ for each $T \subseteq \binom{[n]}{2}$ (the "Fourier coefficients"). These Fourier coefficients will be fixed by our insistence on the pseudoexpectation being pseudocalibrated with respect to the planted distribution $G(n, 1/2, \omega)$.

Definition 5.4.1 ($\tilde{\mathbb{E}}$ of degree d, clique-size ω, truncation τ)**.** Let $S \subseteq [n]$ be a set of vertices

of size $|S| \leqslant d$. Let $T \subseteq \binom{[n]}{2}$ be a set of edges. Let $\chi_T = \prod_{e \in T} G_e$. Let

$$\widehat{\tilde{\mathbb{E}}[x_S]}(T) = \begin{cases} \mathbb{E}_{(G,x)\sim G(n,1/2,\omega)}[\chi_T(G) x_S] & \text{if } |\mathcal{V}(T) \cup S| \leqslant \tau \\ 0 & \text{otherwise}. \end{cases}$$

As usual, $\tilde{\mathbb{E}}[x_S] = \sum_{T \subseteq \binom{[n]}{2}} \widehat{\tilde{\mathbb{E}}[x_S]}(T) \cdot \chi_T(G)$.

The Fourier coefficients can in fact be explicitly computed easily:

Lemma 5.4.2. *Let $T \subseteq \binom{[n]}{2}$, $S \subseteq [n]$ and $\mathcal{V}(T) \subseteq [n]$ be the vertices incident to edges in T. Then*

$$\mathop{\mathbb{E}}_{(H,x)\sim G(n,1/2,\omega)}[\chi_T \cdot x_S] = \left(\tfrac{\omega}{n}\right)^{|\mathcal{V}(T)\cup S|}.$$

Proof. Throughout this proof, we suppress explicit notation for the underlying random variable which is $(H, x) \sim G(n, \frac{1}{2}, \omega)$. We claim that $\mathbb{E}[\chi_T \cdot x_S] = \mathbb{P}[x_{\mathcal{V}(T)\cup S} = 1]$. To see this, note that

$$\begin{aligned} \mathbb{E}[\chi_T \cdot x_S] = \mathbb{P}[x_{\mathcal{V}(T)\cup S} = 1] \cdot \mathbb{E}[\chi_T \cdot x_S \mid x_{\mathcal{V}(T)\cup S} = 0] \\ + (1 - \mathbb{P}[x_{\mathcal{V}(T)\cup S} = 1]) \cdot \mathbb{E}[\chi_T \cdot x_S \mid x_{\mathcal{V}(T)\cup S} = 0]. \quad (5.4.1) \end{aligned}$$

We note that the second term above is 0. It's easy to see if $x_S = 0$. Otherwise, $x_{\mathcal{V}(T)} = 0$, and there is an edge $e \in T$ but not contained in the clique x. Thus,

$$\mathbb{E}[\chi_e \chi_{T\setminus e} \cdot x_S \mid x_{\mathcal{V}(T)\cup S} = 0] = 0.$$

If $x_{\mathcal{V}(T)\cup S} = 1$ then $\chi_T = 1$, so $\mathbb{E}[\chi_T \cdot x_S \mid x_{\mathcal{V}(T)\cup S} = 1] = 1$. By a simple computation,

$$\mathbb{P}[x_{V(T)\cup S} = 1] = \left(\tfrac{\omega}{n}\right)^{|\mathcal{V}(T)\cup S|}. \qquad \square$$

As discussed in Section 5.1.3, our construction of $\tilde{\mathbb{E}}$ is pseudocalibrated. The following lemma captures this formally. We include the (straightforward) proof in Appendix A.0.1.

Lemma 5.4.3. *Let $f_G(x) = \sum_{|S|\leqslant 2d} c_S(G) \cdot x_S$ be a real-valued polynomial on $\{0,1\}^n$ whose coefficients have degree at most τ when expressed in the ± 1 indicators G_e for edges in G. Then, $\mathbb{E}_{G\sim G(n,\frac{1}{2})}[\tilde{\mathbb{E}}[f_G(x)]] = \mathbb{E}_{(H,x)\sim G(n,1/2,\omega)}[f_H(x)]$.*

5.4.2 $\tilde{\mathbb{E}}$ Satisfies Constraints

We now show that the $\tilde{\mathbb{E}}$ defined in the previous section satisfies all linear constraints among (1) – (6) in Section 5.1.2 and has an objective value of ω. That is, 1) $\tilde{\mathbb{E}}[1] \approx 1$, 2) $\tilde{\mathbb{E}}[\sum_{i\in[n]} x_i] \approx \omega$, and 3) $\tilde{\mathbb{E}}[x_S] = 0$ for every $S \subseteq [n]$ which is not a clique in G.

We analyze $\tilde{\mathbb{E}}[1]$ and $\tilde{\mathbb{E}}[\sum_{i\in[n]} x_i]$ in the next lemma and include a proof based on moment-method in Appendix A.0.2.

Lemma 5.4.4. *With high probability, $\tilde{\mathbb{E}}[1] = 1 \pm n^{-\Omega(\varepsilon)}$ and $\tilde{\mathbb{E}}[\sum_{i\in[n]} x_i] = \omega \cdot (1 \pm n^{-\Omega(\varepsilon)})$.*

The next lemma shows that $\tilde{\mathbb{E}}[x_S] = 0$.

Lemma 5.4.5. *With probability 1, if $S \subseteq [n]$ of size at most d is not a clique in G, then $\tilde{\mathbb{E}}[x_S] = 0$.*

Proof. Let $S \subseteq [n]$ have size at most d. Recall that $\mathbb{1}_{S \text{ is a clique in } G} = 2^{-\binom{|S|}{2}} \sum_{T\subseteq\binom{S}{2}} \chi_T$. Becasue the Fourier expansion of $\tilde{\mathbb{E}}[x_S]$ is truncated using the threshold $|\mathcal{V}(T) \cup S| \leqslant \tau$, two Fourier characters $\chi_T, \chi_{T'}$ have the same coefficient in $\tilde{\mathbb{E}}[x_S]$ if $T \oplus T' \subseteq \binom{S}{2}$. So we can factor $\tilde{\mathbb{E}}[x_S] = \mathbb{1}_{S \text{ is a clique in } G} \cdot f_S(G)$ for some function f_S. □

5.4.3 Proof of Main Theorem

Our main technical claim is that $\tilde{\mathbb{E}} = \tilde{\mathbb{E}}_G$ is (approximately) PSD. That is:

Lemma 5.4.6. *With high probability over G from $G(n, 1/2)$, every $p \in \mathbb{P}_d$ satisfies,*

$$\tilde{\mathop{\mathbb{E}}_{G}}[p(x)^2] \geqslant 0$$

It is easy to complete the proof of Theorem 3.0.2 now:

Proof of Theorem 3.0.2. By Lemma 5.4.4, Lemma 5.4.5, and Lemma 5.4.6, there is a universal C so that if $Cd/\varepsilon \leqslant \tau \leqslant (1/C)\varepsilon \log n$, (by a union bound) with high probability the following all hold:

1. $\tilde{\mathbb{E}}[1] = 1 \pm n^{-\Omega(\varepsilon)}$.
2. $\tilde{\mathbb{E}}[x_S] = 0$ for every S of size at most d not a clique in G.
3. $\tilde{\mathbb{E}}[\sum_i x_i] \geqslant (1 - n^{-\Omega(\varepsilon)})\omega$.
4. $\tilde{\mathbb{E}}[p(x)^2] \geqslant 0$ for every $p \in \mathbb{P}_d$.

Thus, choose $\varepsilon = (C^2 d/\log n)^{1/2}$ and $\tau = (1/C)\varepsilon \log n$. The operator given by $\tilde{\mathbb{E}}^*[p(x)] = \tilde{\mathbb{E}}[p(x)]/\tilde{\mathbb{E}}[1]$ is a valid degree-d pseudo-distribution with $\tilde{\mathbb{E}}[\sum_i x_i] \geqslant \Omega(n^{1/2-\Theta(d/\log n)^{1/2}})$ as desired.

5.4.4 Proof Plan

As is standard, we can reduce Lemma 5.4.6 to showing that the associated *moment matrix*, is positive semidefinite.

Definition 5.4.7 (Moment Matrix)**.** Let $\mathcal{M} \in \mathbb{R}^{\binom{[n]}{\leq d} \times \binom{[n]}{\leq d}}$ be given by $\mathcal{M}(I, J) = \tilde{\mathbb{E}}[x_I x_J]$.

Thus, Lemma 5.4.6 is equivalent to showing:

Lemma 5.4.8. *With high probability,* $\mathcal{M} \succeq 0$.

At a high level our plan involves first getting an approximate factorization of the moment matrix $\mathcal{M} = \mathcal{L}\mathcal{Q}_0\mathcal{L}^\intercal + "error"$ for appropriately defined matrices $\mathcal{L}$ and $\mathcal{Q}_0$. This step is the key technical part of the proof - given such a factorization, our task reduces to showing that $\mathcal{Q}_0$ and $\mathcal{L}\mathcal{L}^\intercal$ has large enough positive eigenvalues to compensate for the error. The first approximate factorization step will occupy us in Section 5.5. The technical work in second step involves showing upper bounds on the spectral norms of appropriately defined pieces of $\mathcal{Q}_0$ and is the content of Section 5.6.

□

5.5 Approximate Factorization of the Moment Matrix

5.5.1 Ribbons and Vertex Separators

In this section we get set up for the first step in the proof of Lemma 5.4.8 by setting up some definitions. *Ribbons* will play a crucial role in our analysis:

Definition 5.5.1 (Ribbon)**.** An (I, J)-ribbon $\mathcal{R}$ is a graph with edge set $W_\mathcal{R} \subseteq \binom{[n]}{2}$ and vertex set $V_\mathcal{R} \supseteq \mathcal{V}(W_\mathcal{R}) \cup I \cup J$, for two specially identified subsets $I, J \subseteq [n]$, each of size at most d, called the *left* and the *right ends*, respectively. We sometimes write $\mathcal{V}(\mathcal{R}) \stackrel{\text{def}}{=} V_\mathcal{R}$ and call $|\mathcal{V}(\mathcal{R})|$ the *size* of $\mathcal{R}$. Also, we write $\chi_\mathcal{R}$ for the monomial $\chi_{W_\mathcal{R}}$ where $W_\mathcal{R}$ is the edge set of the ribbon $\mathcal{R}$.

In our analysis, (I, J)-ribbons arise as the terms in the Fourier decomposition of the entry $\mathcal{M}(I, J)$ in the moment matrix. It is important to emphasize that the subsets I and J in an (I, J)-ribbon are allowed to intersect. Also $\mathcal{V}(\mathcal{R})$ can contain vertices that are not in $\mathcal{V}(W_{\mathcal{R}})$ if there are isolated vertices in the ribbon.

Ultimately, we will want to partition a ribbon into three subribbons in such a way that we can express the moment matrix as the sum of positive semidefinite matrices, and some error terms. Our partitioning will be based on minimum vertex separators.

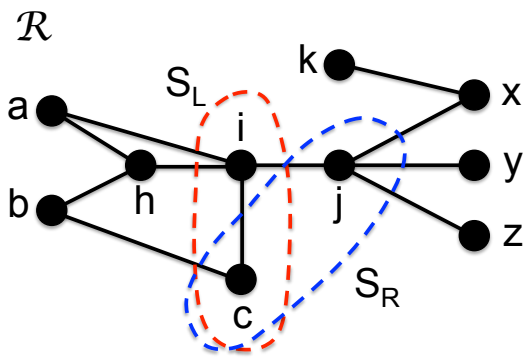
Definition 5.5.2 (Vertex Separator)**.** For an (I, J)-ribbon $\mathcal{R}$ with edge set $W_{\mathcal{R}}$, a subset $Q \subseteq \mathcal{V}(\mathcal{R})$ of vertices is a *vertex separator* if Q separates I and J in $W_{\mathcal{R}}$. A vertex separator is *minimum* if there are no other vertex separators with strictly fewer vertices. The *separator size* of $\mathcal{R}$ is the cardinality of any minimum vertex separator of $\mathcal{R}$.

The following elementary lemma establishes that a ribbon has a unique *leftmost* and *rightmost* vertex separator of minimum size. We defer its proof to Appendix A.0.3.

Lemma 5.5.3 (Leftmost/Rightmost Vertex Separator)**.** *Let $\mathcal{R}$ be an (I, J)-ribbon. There is a unique minimum vertex separator S of $\mathcal{R}$ such that S separates I and Q for any vertex separator Q of $\mathcal{R}$. We call S the* leftmost *separator in $\mathcal{R}$. We define the* rightmost *separator analogously and we denote them by $S_L(\mathcal{R})$ and $S_R(\mathcal{R})$ respectively.*

We illustrate the notion of a leftmost and rightmost vertex separator in the example below.

Let $I = \{a, b, c\}$ and let $J = \{c, x, y, z\}$. The maximum number of vertex disjoint paths from I to J is 2 — for example, we could take the path $\{c\}$ and the path $\{b, h, i, j, z\}$.



The leftmost and rightmost separators are $S_L = \{c, i\}$ and $S_R = \{c, j\}$ respectively. This example illustrates an important point that when I and J intersect, S_L and S_R must both contain $I \cap J$.

5.5.2 Factorization of Monomials

Our factorization of $\mathcal{M}$ will rely on an iterative argument for grouping and factoring the Fourier characters in the decomposition of $\mathcal{M}(I, J)$.

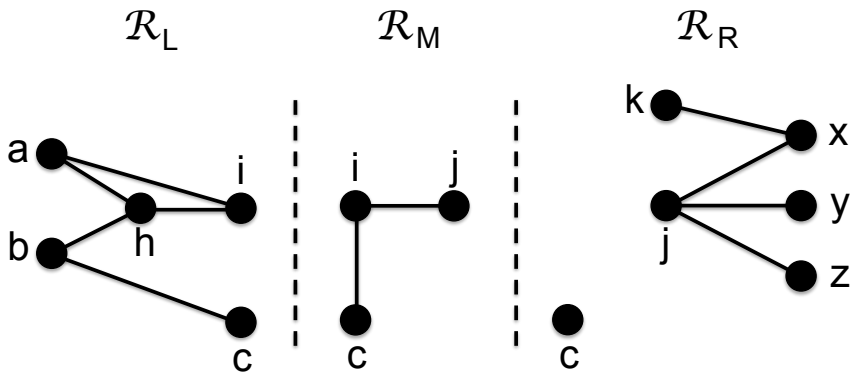
Definition 5.5.4 (Canonical Factorization)**.** Let $\mathcal{R}$ be an (I, J)-ribbon with edge set $W_{\mathcal{R}}$ and vertex set $V_{\mathcal{R}}$. Let V_ℓ be the vertices reachable from I without passing through $S_L(\mathcal{R})$, and similarly for V_r, and let $V_m = V_{\mathcal{R}} \setminus (V_\ell \cup V_r)$. Let $W_\ell \subseteq W_{\mathcal{R}}$ be given by

$$W_\ell = \{(u, v) \in W_{\mathcal{R}} \,:\, u \in V_\ell \text{ and } v \in V_\ell \cup S_L\}$$

and similarly for W_r. Finally, let $W_m = W_{\mathcal{R}} \setminus (W_\ell \cup W_r)$.

Let $\mathcal{R}_\ell$ be the $(I, S_L(\mathcal{R}))$-ribbon with vertex set $V_\ell \cup S_L(\mathcal{R})$ and edge set W_ℓ and similarly for $\mathcal{R}_r$. Let $\mathcal{R}_m$ be the $(S_L(\mathcal{R}), S_R(\mathcal{R}))$-ribbon with vertex set V_m and edge set W_m . The triple $(\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r)$ is the *canonical factorization* of $\mathcal{R}$.

Some facts about the canonical factorization are worth emphasizing. First, W_ℓ, W_m and W_r are disjoint and are a partition of $W_{\mathcal{R}}$ by construction. Hence $\chi_{\mathcal{R}} = \chi_{W_\ell} \cdot \chi_{W_m} \cdot \chi_{W_r}$. Second, some vertices in I may not be in V_ℓ at all. However any such vertices that are in I but not V_ℓ are necessarily in S_L and thus will be contained in $\mathcal{R}_\ell$ anyways. This is why we can say that $\mathcal{R}_\ell$ is an $(I, S_L(\mathcal{R}))$-ribbon. The following illustrates what the canonical factorization would look like in our earlier example:



We chose this example to illustrate a subtle point. The edge (i, c) has both its endpoints in both $\mathcal{R}_\ell$ and $\mathcal{R}_m$. We could in principle choose to place it in either, but we have adopted the convention that because both of its endpoints are in S_L we place it in $\mathcal{R}_m$. In this way, there are no edges within S_L in $\mathcal{R}_\ell$ or within S_R in $\mathcal{R}_m$. Finally, note that there can be isolated vertices in $\mathcal{R}_\ell$ or $\mathcal{R}_r$ but such vertices need to be in I or J respectively.

With the definition of the canonical factorization in hand, we will collect some important properties about it that we will make use of later:

Claim 5.5.5. Let $\mathcal{R}$ be an (I, J)-ribbon with canonical factorization $(\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r)$. Then

$$|\mathcal{V}(\mathcal{R})| = |\mathcal{V}(\mathcal{R}_\ell)| + |\mathcal{V}(\mathcal{R}_m)| + |\mathcal{V}(\mathcal{R}_r)| - |S_L(\mathcal{R})| - |S_R(\mathcal{R})|.$$

Proof. It is important to note that $S_L(\mathcal{R})$ and $S_R(\mathcal{R})$ are not necessarily disjoint (indeed, this happens in the example above). Nevertheless, we know that by construction V_ℓ, V_m and V_r are disjoint and that $S_L(\mathcal{R}) \cup S_R(\mathcal{R}) \subseteq V_m$. Every vertex that appears just once in $S_L(\mathcal{R})$ and $S_R(\mathcal{R})$ appears twice in the canonical factorization. And every vertex that is in $S_L(\mathcal{R}) \cap S_R(\mathcal{R})$ appears three times. Thus

$$|\mathcal{V}(\mathcal{R})| = |\mathcal{V}(\mathcal{R}_\ell)| + |\mathcal{V}(\mathcal{R}_m)| + |\mathcal{V}(\mathcal{R}_r)| - |S_L(\mathcal{R})/S_R(\mathcal{R})| - |S_R(\mathcal{R})/S_L(\mathcal{R})| - 2|S_L(\mathcal{R}) \cap S_R(\mathcal{R})|$$

which completes the proof. □

In the discussion above, we established some properties that a canonical factorization must satisfy. Next we show the reverse direction, that any collection of ribbons that satisfies the below properties must be a canonical factorization. Consider a collection of ribbons $\mathcal{R}_0, \mathcal{R}_1, \mathcal{R}_2$, and the following list of properties:

S_ℓ, S_r **Factorization Conditions for** $\mathcal{R}_0, \mathcal{R}_1, \mathcal{R}_2$ **(Here** $S_\ell, S_r \subseteq [n]$**.).**

1. $\mathcal{R}_0$ is an (I, S_ℓ)-ribbon with $S_L(\mathcal{R}_0) = S_R(\mathcal{R}_0) = S_\ell$, and all vertices in $\mathcal{V}(\mathcal{R}_0)$ are either reachable from I without passing through S_ℓ or are in I or S_ℓ. Finally, $\mathcal{R}_0$ has no edges between vertices in S_ℓ.
2. $\mathcal{R}_2$ is an (S_r, J)-ribbon with $S_L(\mathcal{R}_2) = S_R(\mathcal{R}_2) = S_r$, and all vertices in $\mathcal{V}(\mathcal{R}_2)$ are either reachable from J without passing through S_r or are in J or S_r. Finally, $\mathcal{R}_2$ has no edges between vertices in S_r.
3. $\mathcal{R}_1$ is an (S_ℓ, S_r)-ribbon with $S_L(\mathcal{R}_1) = S_\ell$ and $S_R(\mathcal{R}_1) = S_r$. Every vertex in $\mathcal{V}(\mathcal{R}_1) \setminus (S_\ell \cup S_r)$ has degree at least 1.
4. $W_{\mathcal{R}_0}, W_{\mathcal{R}_1}, W_{\mathcal{R}_2}$ are pairwise disjoint. Also, $V_{\mathcal{R}_0} \cap V_{\mathcal{R}_1} = S_\ell$, $V_{\mathcal{R}_1} \cap V_{\mathcal{R}_2} = S_r$, and $V_{\mathcal{R}_0} \cap V_{\mathcal{R}_2} = S_\ell \cap S_r$.

Lemma 5.5.6. *Let $\mathcal{R}_0, \mathcal{R}_1, \mathcal{R}_2$ be ribbons. Then $(\mathcal{R}_0, \mathcal{R}_1, \mathcal{R}_2)$ is the canonical factorization of the (I, J)-ribbon $\mathcal{R}$ with edge set $W_{\mathcal{R}_0} \oplus W_{\mathcal{R}_1} \oplus W_{\mathcal{R}_2}$ and vertex set $\mathcal{V}(\mathcal{R}_0) \cup \mathcal{V}(\mathcal{R}_1) \cup \mathcal{V}(\mathcal{R}_2)$ if and only if the S_ℓ, S_r factorization conditions hold for $\mathcal{R}_0, \mathcal{R}_1, \mathcal{R}_2$ for some $S_\ell, S_r \subseteq [n]$.*

Proof. If $\mathcal{R}$ is a ribbon with leftmost and rightmost vertex separators S_ℓ and S_r and canonical factorization $(\mathcal{R}_0, \mathcal{R}_1, \mathcal{R}_2)$, then many of the conditions above are automatically satisfied. By construction, $W_{\mathcal{R}_0}, W_{\mathcal{R}_1}, W_{\mathcal{R}_2}$ are pairwise disjoint. Because any edge with both endpoints in S_ℓ is included in $\mathcal{R}_m$ we have that there are no edges between vertices in S_ℓ in $\mathcal{R}_0$, and similarly for $\mathcal{R}_2$. Finally suppose there is a vertex u in $\mathcal{R}_0$. If u is not reachable from I without passing through S_ℓ and is not in I or S_ℓ then it would not be included in $\mathcal{R}_0$. An identical argument holds for $\mathcal{R}_2$.

All that remains is to verify that $S_L(\mathcal{R}_0) = S_R(\mathcal{R}_0) = S_\ell$ and similarly for $\mathcal{R}_1, \mathcal{R}_2$. If $S_\ell = S_L(\mathcal{R})$ is not a minimum-size vertex separator for $\mathcal{R}_0$, then it is also not a minimum-size vertex separator for $\mathcal{R}$, which is impossible. Similarly, if it is not the leftmost separator for $\mathcal{R}_0$ then it was not the leftmost separator for $\mathcal{R}$. Since $\mathcal{R}_0$ is an (I, S_ℓ)-ribbon and S_ℓ is a minimum-size separator, it must also be the right-most minimum-size separator.

Now in the reverse direction, suppose that $\mathcal{R}_0, \mathcal{R}_1, \mathcal{R}_2$ are ribbons that meet the S_ℓ, S_r factorization conditions. We claim that S_ℓ is the leftmost separator for $\mathcal{R}$. If not, then either their is a smaller vertex separator, or there is a vertex separator S'_ℓ of the same size that separates I and S_ℓ. To rule out the former case, note that since S_ℓ and S_r are both minimum vertex separators for $\mathcal{R}_1$, we must have $|S_\ell| = |S_r|$. Then it follows from the S_ℓ, S_r factorization conditions that there are $|S_\ell|$ vertex disjoint paths from I to J, but this would contradict the fact that there is a vertex separator with fewer than $|S_\ell|$ vertices.

In the latter case, any other vertex separator S'_ℓ of the same size that separates I and S_ℓ would contradict the condition $S_L(\mathcal{R}_0) = S_\ell$. An identical argument shows that S_r is the rightmost separator for $\mathcal{R}$.

Finally, by assumption all the vertices in $\mathcal{V}(\mathcal{R}_0)$ are either reachable from I without passing through S_ℓ or are in I or S_ℓ and hence would be included in $\mathcal{R}_0$. Similarly, there are no edges in $W_{\mathcal{R}_0}$ with both endpoints in S_ℓ. Thus if we were to compute the canonical factorization for $\mathcal{R}$ we would get the same set of vertices in each ribbon and the same partition of the edges. □

5.5.3 Factorization of Matrix Entries

This leads to our first factorization of the entries $\mathcal{M}(I, J)$ of $\mathcal{M}$. Unfortunately, the error terms in this first attempt will be too large. Using canonical factorizations and Claim 5.5.5, for any $I, J \subseteq [n]$ of size at most d we can write

$$\begin{aligned}\mathcal{M}(I,J) &= \sum_{\substack{\mathcal{R} \text{ an } (I,J)\text{-ribbon with edge set } W,\\ |\mathcal{V}(W)|\leqslant\tau\\ \text{canonical factorization } (\mathcal{R}_\ell,\mathcal{R}_m,\mathcal{R}_r)}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R})|} \cdot \chi_{\mathcal{R}_\ell} \cdot \chi_{\mathcal{R}_m} \cdot \chi_{\mathcal{R}_r} \\ &= \sum_{\substack{S_\ell,S_r\subseteq[n]\\ |S_\ell|=|S_r|\leqslant d}} \left(\frac{\omega}{n}\right)^{-\frac{|S_\ell|+|S_r|}{2}} \\ &\qquad \sum_{\substack{\mathcal{R}_\ell,\mathcal{R}_m,\mathcal{R}_r\subseteq\binom{[n]}{2}\\ \text{satisfying } S_\ell,S_r \text{ factorization conditions}\\ \text{and } |\mathcal{V}(\mathcal{R}_\ell)\cup\mathcal{V}(\mathcal{R}_m)\cup\mathcal{V}(\mathcal{R}_r)|\leqslant\tau}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_\ell)|+|\mathcal{V}(\mathcal{R}_m)|+|\mathcal{V}(\mathcal{R}_r)|-\frac{|S_\ell|+|S_r|}{2}} \cdot \chi_{\mathcal{R}_\ell} \cdot \chi_{\mathcal{R}_m} \cdot \chi_{\mathcal{R}_r}\end{aligned} \tag{5.5.1}$$

Notice that except for the disjointness condition, the S_ℓ, S_r factorization conditions can be separated into condition 1 for $\mathcal{R}_\ell$, condition 3 for $\mathcal{R}_m$, and condition 2 for $\mathcal{R}_r$. We use this

to rewrite as:

$$= \sum_{\substack{S_\ell, S_r \subseteq [n] \\ |S_\ell| = |S_r| \leqslant d}} \left(\frac{\omega}{n}\right)^{-\frac{|S_\ell|+|S_r|}{2}} \left(\sum_{\substack{\mathcal{R}_\ell \text{ having 1} \\ |\mathcal{V}(\mathcal{R}_\ell)| \leqslant \tau}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_\ell)|} \chi_{\mathcal{R}_\ell} \right) \cdot \tag{5.5.2}$$

$$\left(\sum_{\substack{\mathcal{R}_m \text{having 3} \\ |\mathcal{V}(\mathcal{R}_m)| \leqslant \tau}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_m)| - \frac{|S_\ell|+|S_r|}{2}} \chi_{\mathcal{R}_m} \right) \left(\sum_{\substack{\mathcal{R}_r \text{having 2} \\ |\mathcal{V}(\mathcal{R}_r)| \leqslant \tau}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_r)|} \chi_{\mathcal{R}_r} \right) \tag{5.5.3}$$

$$\underbrace{- \sum_{\substack{S_\ell, S_r \subseteq [n] \\ |S_\ell| = |S_r| \leqslant d}} \left(\frac{\omega}{n}\right)^{-\frac{|S_\ell|-|S_r|}{2}} \sum_{\substack{\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r \\ \text{satisfying } S_\ell, S_r \text{ conditions} \\ |\mathcal{V}(\mathcal{R}_\ell)|, |\mathcal{V}(\mathcal{R}_m)|, |\mathcal{V}(\mathcal{R}_r)| \leqslant \tau, \\ |\mathcal{V}(\mathcal{R}_\ell) \cup \mathcal{V}(\mathcal{R}_m) \cup \mathcal{V}(\mathcal{R}_r)| > \tau}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_\ell)|+|\mathcal{V}(\mathcal{R}_m)|+|\mathcal{V}(\mathcal{R}_r)| - \frac{|S_\ell|+|S_r|}{2}} \cdot \chi_{\mathcal{R}_\ell} \cdot \chi_{\mathcal{R}_m} \cdot \chi_{\mathcal{R}_r}}_{\stackrel{\text{def}}{=} \xi_0(I,J), \text{ the error from ribbon size}} \tag{5.5.4}$$

$$\underbrace{- \sum_{\substack{S_\ell, S_r \subseteq [n] \\ |S_\ell| = |S_r| \leqslant d}} \left(\frac{\omega}{n}\right)^{-\frac{|S_\ell|-|S_r|}{2}} \sum_{\substack{\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r \text{ satisfying} \\ 1,3,2 \text{ and not } 4 \\ |\mathcal{V}(\mathcal{R}_\ell)|, |\mathcal{V}(\mathcal{R}_m)|, |\mathcal{V}(\mathcal{R}_r)| \leqslant \tau}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_\ell)|+|\mathcal{V}(\mathcal{R}_m)|+|\mathcal{V}(\mathcal{R}_r)| - \frac{|S_\ell|+|S_r|}{2}} \cdot \chi_{\mathcal{R}_\ell} \cdot \chi_{\mathcal{R}_m} \cdot \chi_{\mathcal{R}_r}}_{\stackrel{\text{def}}{=} E_0(I,J), \text{ the error from ribbon nondisjointness}} \cdot \tag{5.5.5}$$

5.5.4 Factorization of the Matrix $\mathcal{M}$

In lines 5.5.4 and 5.5.5 we have defined two error matrices, $\xi_0, E_0 \in \mathbb{R}^{\binom{[n]}{\leqslant d} \times \binom{[n]}{\leqslant d}}$. Inspired by the factorization of $\mathcal{M}(I,J)$ in line 5.5.3, we define another pair of matrices as

follows:

$$Q_0 \in \mathbb{R}^{\binom{[n]}{d} \times \binom{[n]}{d}} \quad \text{given by} \quad Q_0(S_\ell, S_r) = \sum_{\substack{\mathcal{R}_m \text{ having } 3 \\ |\mathcal{V}(\mathcal{R}_m)| \leqslant \tau}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_m)| - \frac{|S_\ell|+|S_r|}{2}} \chi_{\mathcal{R}_m}$$

$$\mathcal{L} \in \mathbb{R}^{\binom{[n]}{d} \times \binom{[n]}{d}} \quad \text{given by} \quad \mathcal{L}(I, S) = \left(\frac{\omega}{n}\right)^{-\frac{|S|}{2}} \sum_{\substack{\mathcal{R}_\ell \text{ having } 1 \\ |\mathcal{V}(\mathcal{R}_\ell)| \leqslant \tau}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_\ell)|} \chi_{\mathcal{R}_\ell} .$$

The powers of (ω/n) are split between Q_0 and $\mathcal{L}$ so that the typical of eigenvalue of Q_0 will be approximately 1 (although it will be some time before we are prepared to prove that).

The equation in lines 5.5.3, 5.5.4, and 5.5.5 can be written succinctly as

$$\mathcal{M} = \mathcal{L} Q_0 \mathcal{L}^\top - \xi_0 - E_0 .$$

As we will see later, with high probability $Q_0 \succeq 0$, and thus also $\mathcal{L} Q_0 \mathcal{L}^\top \succeq 0$. So long as τ is sufficiently large, the spectral norm $\|\xi_0\|$ of the error term that accounts for ribbons whose size is too large will be negligible. However, the error E_0 does not turn out to be negligible. To overcome this we will apply a similar factorization approach to E_0 as we did for $\mathcal{M}$; iterating this factorization will push down the error from ribbon nondisjointness.

We record an elementary fact about Q_0:

Lemma 5.5.7. *Let Π be the projector to $\mathsf{Span}\{e_C : C \in C_{\leqslant d}\}$. Then $Q_0 = \Pi Q_0 = Q_0 \Pi$.*

Proof. Suppose S is not a clique in G. We need to show that the row $Q_0(S, \cdot)$ is zero. For every entry $Q_0(S, S')$, notice that the Fourier coefficients $\widehat{Q_0(S, S')}(T) = \widehat{Q_0(S, S')}(T')$ if $T, T' \subseteq \binom{[n]}{2}$ disagree only on edges inside S. (That is, $T \oplus T' \subseteq \binom{S}{2}$.) This means that $Q_0(S, S') = \mathbb{1}_{S \text{ is a clique in } G} \cdot f_{S,S'}(G)$ for some function $f_{S,S'}$. □

5.5.5 Iterative Factorization of E_0

We recall now the definition of the matrix $E_0 \in \mathbb{R}^{\binom{[n]}{\leqslant d} \times \binom{[n]}{\leqslant d}}$.

$$E_0(I,J) = \sum_{\substack{S_\ell, S_r \subseteq [n] \\ |S_\ell| = |S_r| \leqslant d}} \left(\frac{\omega}{n}\right)^{-\frac{|S_\ell|+|S_r|}{2}} \sum_{\substack{\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r \text{ satisfying} \\ 1,3,2 \text{ and not } 4 \\ |\mathcal{V}(\mathcal{R}_\ell)|, |\mathcal{V}(\mathcal{R}_m)|, |\mathcal{V}(\mathcal{R}_r)| \leqslant \tau}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_\ell)|+|\mathcal{V}(\mathcal{R}_r)|+|\mathcal{V}(\mathcal{R}_m)|-\frac{|S_\ell|+|S_r|}{2}} \cdot \chi_{\mathcal{R}_\ell} \cdot \chi_{\mathcal{R}_m} \cdot \chi_{\mathcal{R}_r} \cdot$$

In what follows, we will show how to factor a slightly more general sort of matrix; this factorization will be applicable iteratively, starting with E_0.

5.5.5.1 The matrix $\mathcal{E}_c$ and its factorization

To express the family of matrices we will factor, we introduce a relaxation of our definition of ribbon and a corresponding relaxation 3* of condition 3 of the S_ℓ, S_r factorization conditions.

Definition 5.5.8 (Improper Ribbon)**.** An *improper* (I,J)*-ribbon* $\mathcal{R}$ is an (I,J)-ribbon $\mathcal{R}_0$ together with a set $\mathcal{Z}(\mathcal{R}) \subseteq [n]$ of vertices disjoint from $\mathcal{V}(\mathcal{R}_0)$. (Think of adding the vertices $\mathcal{Z}(\mathcal{R})$ to the ribbon $\mathcal{R}_0$ as degree-0 nodes.) We write $\mathcal{V}(\mathcal{R}) = \mathcal{V}(\mathcal{R}_0) \cup \mathcal{Z}(\mathcal{R})$. When we need to distinguish, we sometimes call ordinary ribbons "proper".

Every ribbon is also an improper ribbon by taking $\mathcal{Z}(\cdot) = \emptyset$, and every improper ribbon has a corresponding ribbon given by deleting its degree-0 vertices.

Relaxed Factorization Condition for ribbon $\mathcal{R}_1$ with $S_\ell, S_r \subseteq [n]$.
3*. $\mathcal{R}_1$ is an improper (S_ℓ, S_r)-ribbon.

Let c be a $\mathbb{R}$-valued function $c(\mathcal{R})$ on (possibly improper) ribbons. Let $\mathcal{E}_c \in \mathbb{R}^{\binom{[n]}{\leqslant d}\times\binom{[n]}{\leqslant d}}$ be given by

$$\mathcal{E}_c(I,J) = \sum_{\substack{S_\ell,S_r\subseteq[n]\\ |S_\ell|,|S_r|\leqslant d}} \left(\frac{\omega}{n}\right)^{-\frac{|S_\ell|+|S_r|}{2}} \tag{5.5.6}$$

$$\sum_{\substack{\mathcal{R}_\ell,\mathcal{R}_m,\mathcal{R}_r \text{ satisfying}\\ 1,3^*,2 \text{ and not } 4\\ |\mathcal{V}(\mathcal{R}_\ell)|,|\mathcal{V}(\mathcal{R}_m)|,|\mathcal{V}(\mathcal{R}_r)|\leqslant\tau}} c(\mathcal{R}_m)\left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_\ell)|+|\mathcal{V}(\mathcal{R}_r)|+|\mathcal{V}(\mathcal{R}_m)|-\frac{|S_\ell|+|S_r|}{2}} \cdot \chi_{\mathcal{R}_\ell}\cdot\chi_{\mathcal{R}_m}\cdot\chi_{\mathcal{R}_r}\cdot \tag{5.5.7}$$

Note that 3 is a strictly more restrictive condition than 3*. Hence we can define the function c_0 by $c_0(\mathcal{R}_m) = 1$ if $\mathcal{R}_m$ satisfies 3 and $c_0(\mathcal{R}_m) = 0$ otherwise. Then $E_0 = \mathcal{E}_{c_0}$. In this subsection, we will show how to factor any matrix of the form $\mathcal{E}_c$ as

$$\mathcal{E}_c = \mathcal{L}Q_{c'}\mathcal{L}^\top - \mathcal{E}_{c'} - \xi_c$$

for some function c' on ribbons and matrices $Q_{c'}, \xi_c \in \mathbb{R}^{\binom{[n]}{\leqslant d}\times\binom{[n]}{\leqslant d}}$ where $\|\xi_c\|$ is negligible with high probability.

Just as our initial factorization of $\mathcal{M}$ began with a factorization of each ribbon appearing in the Fourier expansion, our factorization of $\mathcal{E}_c$ depends on a factorization for each triple $(\mathcal{R}_\ell,\mathcal{R}_m,\mathcal{R}_r)$ appearing in 5.5.7. Since they do not satisfy 4, there must be some vertices occurring in more than one of $\mathcal{V}(\mathcal{R}_\ell),\mathcal{V}(\mathcal{R}_m),\mathcal{V}(\mathcal{R}_\ell)$. Before, the canonical factorization depended on the leftmost and rightmost vertex separators in an (I,J)-ribbon $\mathcal{R}$ separating I from J. But now we will be interested in leftmost and rightmost separators that separate both I and J from each other and from these repeated vertices.

Definition 5.5.9 (Separating Factorization)**.** Let $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r$ be ribbons satisfying S_ℓ, S_r factorization conditions 1, 3*, 2 but not 4, with $|\mathcal{V}(\mathcal{R}_\ell)|, |\mathcal{V}(\mathcal{R}_m)|, |\mathcal{V}(\mathcal{R}_r)| \leqslant \tau$. Let $\mathcal{R}$ be the (I, J)-ribbon with edge set $W_{\mathcal{R}_\ell} \oplus W_{\mathcal{R}_m} \oplus W_{\mathcal{R}_r}$ and vertex set $\mathcal{V}(\mathcal{R}_\ell) \cup \mathcal{V}(\mathcal{R}_m) \cup \mathcal{V}(\mathcal{R}_r)$. (Thus, $\chi_{\mathcal{R}_\ell} \cdot \chi_{\mathcal{R}_m} \cdot \chi_{\mathcal{R}_r} = \chi_{\mathcal{R}}$.)

Let S'_ℓ be the leftmost minimum-size vertex separator in $\mathcal{R}$ which separates I from J and any vertices appearing in more than one of $\mathcal{V}(\mathcal{R}_\ell), \mathcal{V}(\mathcal{R}_m), \mathcal{V}(\mathcal{R}_r)$. Similarly, let S'_r be the rightmost minimum-size vertex separator in $\mathcal{R}$ separating J from I and these repeated vertices. (Notice that S'_ℓ and S'_r could have different sizes.)

Let V'_ℓ be the vertices reachable from I without passing through S'_ℓ and similarly for V'_r. Let $V'_m = V_{\mathcal{R}} \setminus (V'_\ell \cup V'_r)$. Let $W'_\ell = \{(u, v) \in W_{\mathcal{R}} : u \in V_\ell, v \in V_\ell \cup S'_\ell\}$ and similarly for W'_r, and let $W'_m = W_{\mathcal{R}} \setminus (W'_\ell \cup W'_r)$.

Let $\mathcal{R}'_\ell$ be the (I, S'_ℓ)-ribbon with vertex set $V'_\ell \cup S'_\ell$ and edge set W'_ℓ and let $\mathcal{R}'_r$ be the (S'_r, J)-ribbon with vertex set $V'_r \cup S'_r$ and edge set W'_r. Finally, let $\mathcal{R}'_m$ be the improper (S'_ℓ, S'_r)-ribbon with edge set W'_m and vertex set $(\mathcal{V}(\mathcal{R}) \setminus (V'_\ell \cup V'_r)) \cup S'_\ell \cup S'_r)$.

Note that $\chi_{\mathcal{R}_\ell} \cdot \chi_{\mathcal{R}_m} \cdot \chi_{\mathcal{R}_r} = \chi_{\mathcal{R}'_\ell} \cdot \chi_{\mathcal{R}'_m} \cdot \chi_{\mathcal{R}'_r}$ if $\mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r$ is the separating factorization for $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r$. We can use this to rewrite $\mathcal{E}_c$ as

$$\mathcal{E}_c(I, J) = \sum_{\substack{S_\ell, S_r \subseteq [n] \\ |S_\ell|, |S_r| \leqslant d}} \left(\frac{\omega}{n}\right)^{-\frac{|S_\ell|+|S_r|}{2}} \sum_{\substack{\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r \text{ satisfying} \\ 1,3^*,2 \text{ and not } 4 \\ |\mathcal{V}(\mathcal{R}_\ell)|, |\mathcal{V}(\mathcal{R}_m)|, |\mathcal{V}(\mathcal{R}_r)| \leqslant \tau \\ \text{separating factorization} \\ \mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r, S'_\ell, S'_r}} c(\mathcal{R}_m) \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_\ell)|+|\mathcal{V}(\mathcal{R}_r)|+|\mathcal{V}(\mathcal{R}_m)|-\frac{|S_\ell|+|S_r|}{2}} \cdot \chi_{\mathcal{R}'_\ell} \cdot \chi_{\mathcal{R}'_m} \cdot \chi_{\mathcal{R}'_r} \tag{5.5.8}$$

Our goal is to find some coefficient function c' on (improper) ribbons and a matrix $Q_{c'}$ so that this is approximately equal to $\mathcal{L}Q_{c'}\mathcal{L}^\top - \mathcal{E}_{c'}$. For c' yet to be chosen, we take

$$Q_{c'}(S'_\ell, S'_r) \stackrel{\text{def}}{=} \sum_{\substack{\mathcal{R}'_m \text{ having } 3^* \\ |\mathcal{V}(\mathcal{R}'_m)| \leqslant \tau}} c'(\mathcal{R}'_m) \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}'_m)| - \frac{|S'_\ell|+|S'_r|}{2}} \chi_{\mathcal{R}'_m}$$

and have that

$$\mathcal{L}Q_{c'}\mathcal{L}^\top(I,J) - \mathcal{E}_{c'}(I,J) = \sum_{\substack{S'_\ell, S'_r \subseteq [n] \\ |S'_\ell|, |S'_r| \leqslant d}} \left(\frac{\omega}{n}\right)^{-\frac{|S'_\ell|+|S'_r|}{2}} \sum_{\substack{\mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r \text{ satisfying} \\ 1, 3^*, 2, \text{ and } 4 \\ |\mathcal{V}(\mathcal{R}'_\ell)|, |\mathcal{V}(\mathcal{R}'_m)|, |\mathcal{V}(\mathcal{R}'_r)| \leqslant \tau}} c'(\mathcal{R}'_m) \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}'_\ell)|+|\mathcal{V}(\mathcal{R}'_r)|+|\mathcal{V}(\mathcal{R}'_m)| - \frac{|S'_\ell|+|S'_r|}{2}} \cdot \chi_{\mathcal{R}'_\ell} \cdot \chi_{\mathcal{R}'_m} \cdot \chi_{\mathcal{R}'_r} \cdot \tag{5.5.9}$$

We will compare (5.5.8) and (5.5.9) by collecting like terms, but first we handle the discrepancy in the size bounds on the ribbons with a corresponding error term ξ_c. The following matrix is similar to $\mathcal{E}_c$, but places a size bound on the ribbons in the separating factorization $|\mathcal{V}(\mathcal{R}'_\ell)|, |\mathcal{V}(\mathcal{R}'_m)|, |\mathcal{V}(\mathcal{R}'_r)| \leqslant \tau$. We define

$$\mathcal{E}'_c(I,J) = \sum_{\substack{S_\ell, S_r \subseteq [n] \\ |S_\ell|, |S_r| \leqslant d}} \left(\frac{\omega}{n}\right)^{-\frac{|S_\ell|+|S_r|}{2}} \sum_{\substack{\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r \text{ satisfying} \\ 1, 3^*, 2 \text{ and not } 4 \\ \text{separating factorization} \\ \mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r, S'_\ell, S'_r \\ |\mathcal{V}(\mathcal{R}'_\ell)|, |\mathcal{V}(\mathcal{R}'_m)|, |\mathcal{V}(\mathcal{R}'_r)| \leqslant \tau}} c(\mathcal{R}_m) \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_\ell)|+|\mathcal{V}(\mathcal{R}_r)|+|\mathcal{V}(\mathcal{R}_m)| - \frac{|S_\ell|+|S_r|}{2}} \cdot \chi_{\mathcal{R}'_\ell} \cdot \chi_{\mathcal{R}'_m} \cdot \chi_{\mathcal{R}'_r}$$

We take $\xi_c = \mathcal{E}'_c - \mathcal{E}_c$ and we will show below that with high probability the error $\|\xi_c\|$ is negligible. Before doing this, we show that $\mathcal{E}'_c$ is exactly equal to $\mathcal{L}^\top Q_{c'} \mathcal{L}^\top - \mathcal{E}_{c'}$ for the correct choice of c'.

To collect like terms, it helps to define the following quantity $\gamma_{\mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r, I, J, S'_\ell, S'_r}$.

$$\gamma_{\mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r, I, J, S'_\ell, S'_r} \stackrel{\text{def}}{=} \sum_{\substack{\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r \text{ satisfying} \\ \text{1,3*,2 and not 4 for some } S_\ell, S_r \\ \text{separating factorization } \mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r, S'_\ell, S'_r}} c(\mathcal{R}_m) \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_\ell)| + |\mathcal{V}(\mathcal{R}_m)| + |\mathcal{V}(\mathcal{R}_r)| + \frac{|S'_\ell| + |S'_r|}{2} - |S_\ell| - |S_r|}.$$

Then we can rewrite $\mathcal{E}'_c(I, J)$ again as

$$\mathcal{E}'_c(I, J) = \sum_{\substack{S'_\ell, S'_r \subseteq [n] \\ |S'_\ell|, |S'_r| \leqslant d}} \left(\frac{\omega}{n}\right)^{-\frac{|S'_\ell| + |S'_r|}{2}} \sum_{\substack{\mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r \\ \text{satisfying 1, 3*, 2, 4 for } S'_\ell, S'_r \\ |\mathcal{V}(\mathcal{R}'_\ell)|, |\mathcal{V}(\mathcal{R}'_m)|, |\mathcal{V}(\mathcal{R}_r)| \leqslant \tau}} \gamma_{\mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r, I, J, S'_\ell, S'_r} \cdot \mathcal{X}_{\mathcal{R}'_\ell} \cdot \mathcal{X}_{\mathcal{R}'_m} \cdot \mathcal{X}_{\mathcal{R}'_r}$$

We will obtain $\mathcal{E}'_c = \mathcal{L}^\top Q_{c'} \mathcal{L}^\top - \mathcal{E}_{c'}$ if we define $c'(\mathcal{R}'_m)$ so that

$$c'(\mathcal{R}'_m) \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}'_\ell)| + |\mathcal{V}(\mathcal{R}'_r)| + |\mathcal{V}(\mathcal{R}'_m)| - \frac{|S'_\ell| + |S'_r|}{2}} = \gamma_{\mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r, I, J, S'_\ell, S'_r}$$

To express this in terms of the function c, we expand out $\gamma_{\mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r, I, J, S'_\ell, S'_r}$. It is useful to define:

Definition 5.5.10. Let

$$r = (|\mathcal{V}(\mathcal{R}_\ell)| + |\mathcal{V}(\mathcal{R}_m)| + |\mathcal{V}(\mathcal{R}_r)| - |S_\ell| - |S_r|) - (|\mathcal{V}(\mathcal{R}'_\ell)| + |\mathcal{V}(\mathcal{R}'_m)| + |\mathcal{V}(\mathcal{R}'_r)| - |S'_\ell| - |S'_r|).$$

(The ribbons $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r, \mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r$ will always be clear from context.)

Note that $(\mathcal{V}(\mathcal{R}_\ell)| + |\mathcal{V}(\mathcal{R}_m)| + |\mathcal{V}(\mathcal{R}_r)| - |S_\ell| - |S_r|)$ is the total number of vertices we would have in the (I, J)-ribbon with vertex set $\mathcal{V}(\mathcal{R}_\ell) \cup \mathcal{V}(\mathcal{R}_m) \cup \mathcal{V}(\mathcal{R}_\ell)$ if $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r$ satisfied condition 4 (which they do not!). Similarly, $(|\mathcal{V}(\mathcal{R}'_\ell)|+|\mathcal{V}(\mathcal{R}'_m)|+|\mathcal{V}(\mathcal{R}'_r)|-|S'_\ell|-|S'_r|)$ is the total number of vertices in the (I, J)-ribbon with edge set $\mathcal{W}(\mathcal{R}'_\ell) \cup \mathcal{W}(\mathcal{R}'_m) \cup \mathcal{W}(\mathcal{R}'_r)$ and vertex set $\mathcal{V}(\mathcal{R}'_\ell) \cup \mathcal{V}(\mathcal{R}'_m) \cup \mathcal{V}(\mathcal{R}'_r)$. Thus, r is the number of vertices occurring with multiplicity higher than they should in $\mathcal{V}(\mathcal{R}_\ell) \cup \mathcal{V}(\mathcal{R}_m) \cup \mathcal{V}(\mathcal{R}_r)$.

We can rewrite the γ's as

$$\gamma'_{\mathcal{R}'_\ell,\mathcal{R}'_m,\mathcal{R}'_r,I,J,S'_\ell,S'_r} = \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}'_\ell)|+|\mathcal{V}(\mathcal{R}'_m)|+|\mathcal{V}(\mathcal{R}'_r)|-\frac{|S'_\ell|+|S'_r|}{2}} \sum_{\substack{\mathcal{R}_\ell,\mathcal{R}_m,\mathcal{R}_r \text{ satisfying} \\ 1,3^*,2 \text{ and not } 4 \text{ for some } S_\ell,S_r \\ r \text{ intersections outside } S_\ell,S_r \\ \text{separating factorization } \mathcal{R}'_\ell,\mathcal{R}'_m,\mathcal{R}'_r S'_\ell,S'_r}} c(\mathcal{R}_m)\left(\frac{\omega}{n}\right)^r .$$

Thus, we will have that $\mathcal{E}'_c = \mathcal{L}\mathcal{Q}_{c'}\mathcal{L}^\top - \mathcal{E}_{c'}$ if and only if for every (S'_ℓ, S'_r)-ribbon $\mathcal{R}'_m$ and every $\mathcal{R}'_\ell, \mathcal{R}'_r$ satisfying 1, 2,

$$c'(\mathcal{R}'_m) = \sum_{\substack{\mathcal{R}_\ell,\mathcal{R}_m,\mathcal{R}_r \text{ satisfying} \\ 1,3^*,2 \text{ and not } 4 \text{ for some } S_\ell,S_r \\ r \text{ intersections outside } S_\ell,S_r \\ \text{separating factorization } \mathcal{R}'_\ell,\mathcal{R}'_m,\mathcal{R}'_r S'_\ell,S'_r}} c(\mathcal{R}_m)\left(\frac{\omega}{n}\right)^r .$$

Note that for this to happen, the right hand side must be independent of $\mathcal{R}'_\ell$ and $\mathcal{R}'_r$. If this is the case, then we can define

$$c'(\mathcal{R}'_m) \stackrel{\text{def}}{=} \sum_{\substack{\mathcal{R}_\ell,\mathcal{R}_m,\mathcal{R}_r \text{ satisfying} \\ 1,3^*,2 \text{ and not } 4 \text{ for some } S_\ell,S_r \\ r \text{ intersections outside } S_\ell,S_r \\ \text{separating factorization } \mathcal{R}'_\ell,\mathcal{R}'_m,\mathcal{R}'_r S'_\ell,S'_r}} c(\mathcal{R}_m)\left(\frac{\omega}{n}\right)^r \text{ for some } \mathcal{R}'_\ell, \mathcal{R}'_r \text{ satisfying } 1, 2 .$$

The next claim shows that, indeed, the choice of $\mathcal{R}'_\ell, \mathcal{R}'_r$ does not matter. (This would not have been true without passing from $\mathcal{E}_c$ to $\mathcal{E}'_c$.)

Claim 5.5.11. Let $\mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r$ satisfy 1, 3*, 2, 4 for some $S'_\ell, S'_r \subseteq [n]$. Let $\mathcal{R}''_\ell$ and $\mathcal{R}''_r$ also satisfy 1 and 2, respectively, for S'_ℓ, S'_r, respectively. Then

$$\sum_{\substack{\mathcal{R}_\ell,\mathcal{R}_m,\mathcal{R}_r \text{ satisfying} \\ \text{1,3*,2 and not 4 for some } S_\ell,S_r \\ r \text{ intersections outside } S_\ell,S_r \\ \text{separating factorization } \mathcal{R}'_\ell,\mathcal{R}'_m,\mathcal{R}'_r S'_\ell,S'_r}} c(\mathcal{R}_m)\left(\frac{\omega}{n}\right)^r = \sum_{\substack{\mathcal{R}_\ell,\mathcal{R}_m,\mathcal{R}_r \text{ satisfying} \\ \text{1,3*,2 and not 4 for some } S_\ell,S_r \\ r \text{ intersections outside } S_\ell,S_r \\ \text{separating factorization } \mathcal{R}''_\ell,\mathcal{R}'_m,\mathcal{R}''_r S'_\ell,S'_r}} c(\mathcal{R}_m)\left(\frac{\omega}{n}\right)^r .$$

(Notice that the left-hand sum refers to $\mathcal{R}'_\ell, \mathcal{R}'_r$ and the right-hand one to $\mathcal{R}''_\ell, \mathcal{R}''_r$.)

Proof. We prove this by showing that there is an exact match between terms on the left hand side and terms on the right hand side. Consider a term on the left hand side. Note that the part of $\mathcal{R}_\ell$ between I and S'_ℓ must be $\mathcal{R}'_\ell$ while the part of $\mathcal{R}_\ell$ between S'_ℓ and S_ℓ becomes part of $\mathcal{R}'_m$. To shift from $\mathcal{R}'_\ell$ to $\mathcal{R}''_\ell$, we simply replace $\mathcal{R}'_\ell$ by $\mathcal{R}''_\ell$ within $\mathcal{R}_\ell$. Similarly, to shift from $\mathcal{R}'_r$ to $\mathcal{R}''_r$, we simply replace $\mathcal{R}'_r$ by $\mathcal{R}''_r$ within $\mathcal{R}_r$.

To show that this gives an exact match, we need to show that r is unaffected by these shifts. To see that shifting from $\mathcal{R}'_\ell$ to $\mathcal{R}''_\ell$ does not affect r, note that all vertices in $\mathcal{V}(\mathcal{R}'_\ell) \setminus S'_\ell$ or $\mathcal{V}(\mathcal{R}'_\ell) \setminus S'_\ell$ must appear in the corresponding $\mathcal{R}_\ell$ and cannot appear in $\mathcal{R}_m$ or $\mathcal{R}_r$. Thus, these vertices always have multiplicity 1 in $\mathcal{V}(\mathcal{R}_\ell) \cup \mathcal{V}(\mathcal{R}_m) \cup \mathcal{V}(\mathcal{R}_r)$. All other vertices (including the ones in S'_ℓ) may appear in $\mathcal{R}_m$ or $\mathcal{R}_r$ as well as $\mathcal{R}_\ell$ but whether or not they do so is unaffected by the shift so their multiplicities in $\mathcal{V}(\mathcal{R}_\ell) \cup \mathcal{V}(\mathcal{R}_m) \cup \mathcal{V}(\mathcal{R}_r)$ are unaffected by the shift and r remains the same. A similar argument holds for shifting from $\mathcal{R}'_r$ to $\mathcal{R}''_r$ □

Remark 5.5.12. For this argument, it was important to keep track of the isolated vertices in $\mathcal{R}'_m$. If we did not keep track of isolated vertices and instead had them disappear, we could

have a situation where there is a vertex v which appears in $\mathcal{R}_\ell$ and $\mathcal{R}_m$ but disappears from $\mathcal{R}'_m$ and is not in S'_ℓ. Since v is no longer in $\mathcal{R}'_m$, $\mathcal{R}''_\ell$ could contain v. If so, then we cannot shift from $\mathcal{R}'_\ell$ to $\mathcal{R}''_\ell$ as this would create a copy of v to the left of S'_ℓ but v should be to the right of S'_ℓ.

Putting everything together, $\mathcal{E}'_c = \mathcal{L}\mathcal{Q}_{c'}\mathcal{L}^\top - \mathcal{E}_{c'}$. Since we defined $\xi_c = \mathcal{E}'_c - \mathcal{E}_c$, we get that $\mathcal{E}_c = \mathcal{L}\mathcal{Q}_c\mathcal{L}^\top - \mathcal{E}_{c'} - \xi_c$, as needed.

The remaining step will be to show that with high probability, the error term ξ_c has negligible norm, which we will accomplish in Section 5.6.5.

Finally, we record the following easy lemma about separating factorizations, which will be useful in the application of the foregoing to factor $\mathcal{E}_0$.

Lemma 5.5.13. *Suppose $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r$ satisfy conditions 1, 3*, 2, but not 4. Let $\mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r$ be their separating factorization, with separators S'_ℓ, S'_r. Then*

$$\frac{|S'_\ell| + |S'_r|}{2} - \frac{|S_\ell| + |S_r|}{2} \geqslant \frac{1}{2}$$

Proof. We claim that $|S_\ell| + |S_r| + 1 \leqslant |S'_\ell| + |S'_r|$ By the violation of condition 4, we cannot have $S_\ell = S'_\ell$ and $S_r = S'_r$. But since S'_ℓ separates I from S_ℓ in $\mathcal{R}_\ell$ and $\mathcal{R}_\ell$ is an (I, S_ℓ)-ribbon whose rightmost vertex separator is also S_ℓ, if $S_\ell \neq S'_\ell$ then $|S_\ell| < |S'_\ell|$, and similarly for S_r and S'_r. So either $|S_\ell| < |S'_\ell|$ or $|S_r| < |S'_r|$, and since the separator sizes are integers, so the difference must be at least 1 and we are done. □

5.5.5.2 Application to E_0 and $\mathcal{M}$

We are ready to define our recursive factorization of E_0. Recall that $c_0(\mathcal{R}_m) = 1$ if $\mathcal{R}_m$ satisfies 3 and $c_0(\mathcal{R}_m) = 0$ otherwise and $E_0 = \mathcal{E}_{c_0}$. Applying the factorization above to

$\mathcal{E}_{c_0}$ we obtain matrices $\xi_1 = \xi_{c_0}, Q_1$, and $\mathcal{E}_{c_1}$. Then of course we can apply the factorization again to $\mathcal{E}_{c_1}$.

Proceeding inductively, for all $i \in [1, 2d]$ let $\xi_i = \xi_{c_{i-1}}, Q_i$, and $\mathcal{E}_{c_i}$ be the matrices given by applying the factorization to $\mathcal{E}_{c_{i-1}}$ at step i.

Claim 5.5.14.

$$\mathcal{M} = \mathcal{L}(Q_0 - Q_1 + Q_2 - \ldots - Q_{2d-1} + Q_{2d})\mathcal{L}^\top - (\xi_0 - \xi_1 + \xi_2 - \ldots - \xi_{2d-1} + \xi_{2d}).$$

Proof. We have that $\mathcal{M} = \mathcal{L}(Q_0)\mathcal{L}^\top - \mathcal{E}_0 - \xi_0$ and $\mathcal{E}_{i-1} = \mathcal{L}Q_i\mathcal{L}^\top - \mathcal{E}_i - \xi_{c_{i-1}} = \mathcal{L}Q_i\mathcal{L}^\top - \mathcal{E}_i - \xi_i$. We prove the claim by starting with the first formula and appliying the second formula for each $i \in [1, 2d]$. At the end, we are left with an extra term $\mathcal{E}_{2d}$. We must show that $\mathcal{E}_{2d} = 0$.

To see why $\mathcal{E}_{2d} = 0$, note that every time we have a separating factorization $\mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r$ for $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r$, the size of either the left separator or the right separator must increase (see Lemma 5.5.13). However, the size of these separators is always at most d, so the only way we can do this for $2d$ steps is if we started with the empty set as the separators and increased the size of either the left or right separator by 1 each time, but not both. However, this too is impossible as if we start with the empty set as the separators, after the first step both the new left separator and the new right separator must have size at least 1. □

5.6 $\mathcal{M}$ is PSD

In this section we combine the factorization of $\mathcal{M}$ in terms of the matrices $\mathcal{L}, Q_i, \xi_i$ that we obtained in Section 5.5 with estimates on the eigenvalues of the Qs and ξs. The

starting point is the following PSDness claim for Q_0.

Lemma 5.6.1. *Let $D \in \mathbb{R}^{\binom{[n]}{\leqslant d} \times \binom{[n]}{\leqslant d}}$ be the diagonal matrix with $D(S,S) = 2^{\binom{|S|}{2}}/4$ if S is a clique in G and 0 otherwise. With high probability, $Q_0 \succeq D$.*

We also need to bound $\|Q_i\|$ for $i > 0$.

Lemma 5.6.2. *Let $D \in \mathbb{R}^{\binom{[n]}{\leqslant d} \times \binom{[n]}{\leqslant d}}$ be the diagonal matrix with $D(S,S) = 2^{\binom{S}{2}}/4$ if S is a clique and is otherwise zero. With high probability, every Q_i for $i \in [1, 2d]$ satisfies*

$$\frac{-D}{8d} \preceq Q_i \preceq \frac{D}{8d}\,.$$

The preceding lemmas are enough to obtain $Q_0 - \ldots + Q_{2d} \succeq D/2$, but in the end we need to work with the matrix $\mathcal{L}(Q_0 - \ldots + Q_{2d})\mathcal{L}^\top - (\xi_0 - \ldots + \xi_{2d})$. The next two lemmas allow us to make this last step.

Lemma 5.6.3. *With high probability, $\Pi\mathcal{L}\Pi\mathcal{L}^\top\Pi \succeq \Omega(\omega/n)^{d+1} \cdot \Pi$, where as usual Π is the projector to $\mathsf{Span}\{e_C : C \in C_{\leqslant d}\}$.*

Finally, we need a bound on the ξ matrices.

Lemma 5.6.4. *With high probability, $\|\xi_0 - \ldots + \xi_{2d}\| \leqslant n^{-16d}$.*

We can now prove Lemma 5.4.8.

Proof of Lemma 5.4.8. By Claim 5.5.14,

$$M = \mathcal{L}(Q_0 - Q_1 + Q_2 - \ldots - Q_{2d-1} + Q_{2d})\mathcal{L}^\top - (\xi_0 - \xi_1 + \xi_2 - \ldots - \xi_{2d-1} + \xi_{2d})\,.$$

By a union bound, with high probability the conclusions of Lemmas 5.6.1, 5.6.2, 5.6.3, and 5.6.4 all hold. By Lemma 5.6.1 and Lemma 5.6.2,

$$Q_0 - Q_1 + Q_2 - \ldots - Q_{2d-1} + Q_{2d} \succeq \frac{D}{2} \succeq \frac{\Pi}{2}\,.$$

where as usual Π is the projector to $\mathsf{Span} e_C : C \in C_{\leqslant d}$. Thus by Lemma 5.6.3, we obtain $\mathcal{L}(Q_0 - \ldots + Q_{2d})\mathcal{L}^\top \succeq \Omega(\omega/n)^{d+1} \cdot \Pi$. Finally, by Lemma 5.6.4 we have

$$\mathcal{M} = \Pi \cdot \mathcal{M} \cdot \Pi \succeq \Omega\left(\frac{\omega}{n}\right)^{d+1} \cdot \Pi - n^{-16d} \cdot \Pi \succeq 0\,. \qquad \square$$

In the next subsections, we prove the foregoing lemmas.

5.6.1 Ribbons and Spectral Norms

We will require bounds on the spectral norm of certain random matrices. Our random matrices arise out of decompositions of the moment matrix from Definition 5.4.7 and are functions of a graph G on vertex set $[n]$. Our norm bounds will hold for what we call *graphical matrices*, that are are defined to capture the matrices that are invariant under a permutation of the vertices of G and are in fact "minimal" such matrices.

We first define the *shape* of a ribbon that identifies the structure of a ribbon up to relabelling.

Definition 5.6.5 (Shape of a Ribbon)**.** For an (I, J)-ribbon $\mathcal{R}$, consider the graph U on the vertex set $[|\mathcal{V}(\mathcal{R})|]$ whose edges are

$$E(U) = \{(i, j) : \text{ there is an edge in } \mathcal{R} \text{ from the } i\text{-th to the } j\text{-th least element of } \mathcal{V}(\mathcal{R})\}\,.$$

(Here we are considering $\mathcal{V}(\mathcal{R})$ to have the usual ordering inherited from $[n]$.) Also, let U have two distinguished subsets of vertices A and B, where $A = \{i :$

the i-th element of $\mathcal{V}(\mathcal{R})$ is in $I\}$, and similarly for B and J. We call U the *shape* of $\mathcal{R}$ and write $\mathsf{shape}(\mathcal{R}) = U$.

We record some observations on shapes of ribbons.

- If $\mathcal{R}$ is a ribbon (not an improper ribbon), its shape satisfies the condition that every vertex outside $A \cup B$ has degree at least 1.
- If, for example, $\mathcal{R}$ is an (I, J) ribbon where $I \cap J = \{1\}$ (which must be the least element in both I and J), then in order for the (I', J')-ribbon $\mathcal{R}'$ to have the same shape as $\mathcal{R}$ it is necessary that $|I' \cap J'| = 1$. More broadly, specifying the shape of a ribbon in particular specifies the pattern of intersection of its endpoints.
- A matrix $M \in \mathbb{R}^{\binom{n}{\leqslant d} \times \binom{n}{\leqslant d}}$ whose entries are given by $M(I, J) = \sum_{\mathcal{R} \text{ an } (I,J)\text{-ribbon with shape } U} \chi_{\mathcal{R}}$ satisfies the assumptions of Lemma 5.6.8. In the following sections, our main strategy will be to decompose the matrices Q_i into matrices of this form.

We are now ready to define graphical matrices.

Definition 5.6.6 (Graphical Matrices)**.** Let U be a graph on the vertex set $[t]$ with two distinguished sets of vertices $A, B \subseteq [t]$. Let $\mathcal{T}(\mathcal{U})$ be the collection of all I, J ribbons with shape U. The graphical matrix $M \in \mathbb{R}^{\binom{[n]}{|A|} \times \binom{[n]}{|B|}}$ of shape U is defined by

$$M(I, J) = \sum_{\mathcal{R}:\mathcal{R} \text{ is an } (I,J)\text{-ribbon and } \mathsf{shape}(\mathcal{R})=U} \chi_{\mathcal{R}}.$$

Example 5.6.7. When U is a graph on 2 vertices with distinguished sets $\{1\}$ and $\{2\}$ of size 1 each and a single edge connecting vertex 1 and 2, the graphical matrix of shape U is just the standard $\{-1, 1\}$-adjacency matrix of the graph G.

The following lemma will be our main tool. It is in essence due to Medarametla and Potechin [84] and special cases of the bound have been proven and used in [65, 66, 43]. We give a proof in the appendix for completeness.

Lemma 5.6.8. *Let U be a graph on $t \leqslant O(\log n)$ vertices, with two distinguished subsets of vertices A and B, and suppose:*

- *U admits p vertex-disjoint paths from $A \setminus B$ to $B \setminus A$.*
- *$|A \cap B| = r$.*
- *Every vertex outside $A \cup B$ has degree at least 1.*

Let $M = M(G)$ be the graphical matrix with shape U. Then, whp, $\|M\| \leqslant n^{\frac{t-p-r}{2}} \cdot 2^{O(t)} \cdot (\log n)^{O(t-r+p)}$.

Remark 5.6.9. Lemma 5.6.8 can be seen as a generalization of the standard upper bound on the spectral norm of the adjacency matrix. Example 5.6.7 shows how adjacency matrix is a graphical matrix with a shape U on 2 vertices with a single edge connecting them, thus $t = 2$, $r = 0$ and $p = 1$. Lemma 5.6.8 thus shows an upper bound of $\sqrt{n}\,\mathrm{poly}\log(n)$ on the spectral norm of the adjacency matrix which is tight up to a $\mathrm{poly}\log(n)$ factor.

5.6.2 Positivity for Q_0 — Proof of Lemma 5.6.1

In this section we prove Lemma 5.6.1, which we restate here.

Lemma (Restatement of Lemma 5.6.1)**.** *Let $D \in \mathbb{R}^{\binom{[n]}{\leqslant d}\times\binom{[n]}{\leqslant d}}$ be the diagonal matrix with $D(S,S) = 2^{\binom{|S|}{2}}/4$ if S is a clique in G and 0 otherwise. With high probability, $Q_0 \succeq D$.*

Proof of Lemma 5.6.1. To begin, we split Q_0 into its diagonal Q_0^{diag} and its off-diagonal $Q_0^{\text{off-diag}}$ parts.

$$Q_0^{\text{diag}}(S_\ell, S_r) = \begin{cases} Q_0(S_\ell, S_r) \text{ if } S_\ell = S_r \\ 0 \text{ otherwise.} \end{cases} \qquad Q_0^{\text{off-diag}}(S_\ell, S_r) = \begin{cases} Q_0(S_\ell, S_r) \text{ if } S_\ell \neq S_r \\ 0 \text{ otherwise.} \end{cases}$$

Then $Q_0 = Q_0^{\text{diag}} + Q_0^{\text{off-diag}}$. Expanding Q_0^{diag},

$$Q_0^{\text{diag}}(S,S) =$$

$$2^{\binom{|S|}{2}} \cdot \mathbb{1}_{S \text{ is a clique}} \cdot \left(1 + \sum_{\substack{\mathcal{R} \text{ nonempty, having 3} \\ \text{and no edges inside } S \\ |S|<|\mathcal{R}|\leqslant\tau}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R})|-|S|} \cdot \chi_{\mathcal{R}} \right) = 2^{\binom{|S|}{2}} \cdot \mathbb{1}_{S \text{ is a clique}} \cdot (1 \pm n^{-\Omega(\varepsilon)})$$

for all $S \in \binom{[n]}{d}$ with high probability by a similar argument as in Lemma 5.4.4 and a union bound.

Next, we bound $\|Q_0^{\text{off-diag}}\|$ be decomposing it according to ribbon shape. Fix $s, t \leqslant \tau$. Let $U_1^{(s,t)}, \ldots, U_q^{(s,t)}$ be all the graphs on vertex set $[t]$ with two distinguished sets of vertices A, B, both of size s, with $|A \cap B| \leqslant s-1$, and where there are $s - |A \cap B|$ vertex-disjoint paths from $A \setminus B$ to $B \setminus A$. Let $M_i^{(s,t)}$ be given by

$$M_i^{(s,t)}(S_\ell, S_r) = \sum_{\mathcal{R} \text{ an } (S_\ell, S_r)\text{-ribbon with shape } U_i^{(s,t)}} \chi_{\mathcal{R}} \cdot$$

Then

$$Q_0^{\text{off-diag}} = \sum_{\substack{s \leqslant d \\ t \leqslant \tau \\ i \leqslant q}} \left(\frac{\omega}{n}\right)^{t-s} \cdot M_i^{(s,t)} .$$

We can apply Lemma 5.6.8 to conclude that with probability at least $1 - O(n^{-100 \log n})$,

$$\left\| \left(\frac{\omega}{n}\right)^{t-s} \cdot M_i^{(s,t)} \right\| \leqslant \left(\frac{\omega}{n}\right)^{t-s} \cdot n^{\frac{t-s}{2}} \cdot 2^{O(t)} \cdot (\log n)^{O(t-|A\cap B|+|A\setminus B|)} \leqslant n^{-\varepsilon(t-s)} \cdot 2^{O(t)} \cdot (\log n)^{O(t-s)} ,$$

where to conclude the bound on the exponent in $(\log n)^{O(t-|A\cap B|+|A\setminus B|)}$ we have used that $t \geqslant 2s - |A \cap B|$.

Notice that for fixed s and t, there are at most $2^{\binom{t}{2}+O(t)}$ unique shapes $U_1^{(s,t)}, \ldots, U_q^{(s,t)}$. Thus, a union bound followed by the triangle inequality, we obtain that for fixed s and t, with probability at least $1 - O(n^{-99 \log n})$,

$$\left\| \left(\frac{\omega}{n}\right)^{t-s} \sum_{i \leqslant q} M_i^{(s,t)} \right\| \leqslant 2^{\binom{t}{2}+O(t)} \cdot n^{-\varepsilon(t-s)} \cdot 2^{O(t)} \cdot (\log n)^{O(t-s)} .$$

Under our assumptions on the parameters d, τ, and ε, this is at most $2^{\binom{s}{2}}/(100\tau)$. Summing over all $t \leqslant \tau$, for a fixed s we have

$$\left\| \left(\frac{\omega}{n}\right)^{t-s} \sum_{\substack{t \leqslant \tau \\ i \leqslant q}} M_i^{(s,t)} \right\| \leqslant \frac{2^{\binom{s}{2}}}{100} .$$

Notice that the above matrix is exactly the block of $Q_0^{\text{off-diag}}$ corresponding to subsets of size s. Together with our bound on Q_0^{diag}, this proves the lemma. □

5.6.3 Norm Bounds for Q_i — Proof of Lemma 5.6.2

In this section we prove Lemma 5.6.2, restated here.

Lemma (Restatement of Lemma 5.6.2)**.** *Let* $D \in \mathbb{R}^{\binom{[n]}{\leqslant d}\times\binom{[n]}{\leqslant d}}$ *be the diagonal matrix with* $D(S,S) = 2^{\binom{S}{2}}/4$ *if* S *is a clique and is otherwise zero. With high probability, every* Q_i *for* $i \in [1, 2d]$ *satisfies*

$$\frac{-D}{8d} \preceq Q_i \preceq \frac{D}{8d}\,.$$

We will need to bound the coefficients $c_i(\mathcal{R}'_m)$ used to define the matrices Q_i which we set up in Section 5.5.

Lemma 5.6.10. *Let* $c_1, \ldots, c_{2d}$ *be the coefficient functions defined in Section 5.5. For all improper* (S_ℓ, S_r)*-ribbons* $\mathcal{R}_m$ *admitting exactly* p *vertex-disjoint paths from* S_ℓ *to* S_r*, and all* $i \leqslant 2d$*, writing* $s = \frac{|S_\ell|+|S_r|}{2}$*,*

$$c_i(\mathcal{R}_m) \leqslant \left(\frac{\omega}{n}\right)^s \cdot n^{\frac{p-|\mathcal{Z}(\mathcal{R}_m)|-i/2}{2}+\varepsilon s}\,.$$

recalling that $\omega = n^{1/2-\varepsilon}$*. Furthermore, if* $\mathcal{R}_m$ *and* $\mathcal{R}'_m$ *have the same shape, then* $c_i(\mathcal{R}_m) = c_i(\mathcal{R}'_m)$*.*

With this lemma in hand we can prove Lemma 5.6.2.

Proof of Lemma 5.6.2. Fix some $0 < i \leqslant 2d$. We will use Lemma 5.6.8, which requires that we first decompose each Q_i into simpler matrices. First of all, for a proper ribbon $\mathcal{R}_m$, let

$$\tilde{c}_i(\mathcal{R}_m) = \sum_{\mathcal{R}'_m \text{ an improper ribbon whose largest proper subribbon is } \mathcal{R}_m} \left(\frac{\omega}{n}\right)^{|\mathcal{Z}(\mathcal{R}'_m)|} \cdot c_i(\mathcal{R}'_m)\,.$$

Note that we include $\mathcal{R}_m$ itself in this sum as a proper ribbon is also an improper ribbon.

Claim 5.6.11. $\tilde{c}_i(\mathcal{R}_m) \leqslant 2(\omega/n)^s \cdot n^{\frac{p-i/2}{2}+\varepsilon s}$, where p is the number of vertex-disjoint paths from S_ℓ to S_r in $\mathcal{R}_m$.

Proof. Consider all of the improper ribbons $\mathcal{R}'_m$ with k isolated vertices whose largest proper subribbon is $\mathcal{R}_m$. For each such ribbon $\mathcal{R}'_m$, by Lemma 5.6.10, $(\omega/n)^k c_i(\mathcal{R}'_m) \leqslant \left(\frac{\omega}{n}\right)^{k+s} \cdot n^{\frac{p-k-i/2}{2}+\varepsilon s}$. There are at most n^k such improper ribbons. Adding all of their contributions together gives at most

$$\left(\frac{\omega}{\sqrt{n}}\right)^k \left(\frac{\omega}{n}\right)^s \cdot n^{\frac{p-i/2}{2}+\varepsilon s} < 2^{-k}(\omega/n)^s \cdot n^{\frac{p-i/2}{2}+\varepsilon s}$$

Summing this up over all $k \geqslant 0$ gives the result. □

Now fix $s_\ell, s_r \leqslant d$ and $t \leqslant \tau$ and let $U_1^{(s_\ell,s_r,t)}, \ldots, U_q^{(s_\ell,s_r,t)}$ be all graphs on the vertex set $[t]$ with two distinguished subsets of vertices: A of size s_ℓ and B of size s_r. Let

$$\begin{aligned} M_j^{(s_\ell,s_r,t)}(S_\ell, S_r) &= \sum_{\mathcal{R} \text{ is an } (S_\ell,S_r)\text{-ribbon with shape } U_j^{(s_\ell,s_r,t)}} \tilde{c}_i(\mathcal{R}) \cdot \left(\frac{\omega}{n}\right)^{t-s} \cdot \chi_{\mathcal{R}} \\ &= \tilde{c}_i(U_j^{(s_\ell,s_r,t)}) \sum_{\mathcal{R} \text{ is an } (S_\ell,S_r)\text{-ribbon with shape } U_j^{(s_\ell,s_r,t)}} \left(\frac{\omega}{n}\right)^{t-s} \cdot \chi_{\mathcal{R}}\,, \end{aligned}$$

where $s = \frac{s_\ell+s_r}{2}$ and we have used the fact that $\tilde{c}_i(\mathcal{R})$ depends only on the shape of $\mathcal{R}$.

Let $r = |A \cap B|$ where A, B are the distinguished sets of vertices for $U_j^{(s_\ell,s_r,t)}$, and let $\tilde{p}$ be the number of vertex-disjoint paths from $A \setminus B$ to $B \setminus A$, so that $p = r + \tilde{p}$. We can apply Lemma 5.6.8 and our bound on $\tilde{c}_i$ to get that with probability $1 - O(n^{-100\log n})$,

$$\begin{aligned} \left\| M_j^{(s_\ell,s_r,t)} \right\| &\leqslant \left(\frac{\omega}{n}\right)^{t-s} \cdot n^{\frac{\tilde{p}+r-i/2}{2}+\varepsilon s} \cdot n^{\frac{t-\tilde{p}-r}{2}} \cdot 2^{O(t)} \cdot (\log n)^{O(t-r+\tilde{p})} \\ &= n^{-\varepsilon(t-s)-i/4} \cdot 2^{O(t)} \cdot (\log n)^{O(t-r+\tilde{p})} \\ &= n^{-\varepsilon(t-s)-i/4} \cdot 2^{O(t)} \cdot (\log n)^{O(t-s)}\,, \end{aligned}$$

where in the last step we have used that $t \geqslant 2s - r$ and $\tilde{p} \leqslant s - r$.

By inspection,

$$Q_i = \sum_{\substack{s_\ell, s_r \leqslant d \\ t \leqslant \tau \\ j \leqslant q}} M_j^{(s_\ell, s_r, t)}.$$

For a fixed t there are at most $2^{\binom{t}{2}+O(t)}$ choices for U, so $q \leqslant 2^{\binom{t}{2}+O(t)}$. Now we fix s_ℓ, s_r and sum over t to obtain the block of Q_i corresponding to size-s_ℓ and size-s_r subsets. By triangle inequality and a union bound, with probability at least $1 - O(n^{-97 \log n})$,

$$\left\| \sum_{\substack{t \leqslant \tau \\ j \leqslant q}} M_j^{(s_\ell, s_r, t)} \right\| \leqslant 2^{\binom{t}{2}+O(t)} \cdot n^{-\varepsilon(t-s)-i/4} \cdot 2^{O(t)} \cdot (\log n)^{O(t-s)}.$$

From our assumptions on d, τ, and ε, this is at most $2^{\binom{s_\ell}{2}/2+\binom{s_r}{2}/2}/100d^3$.

As usual, let Π be the projector to $\mathsf{Span}\{e_C : C \in C_{\leqslant d}\}$. Note that $\Pi Q_i = Q_i \Pi = Q_i$, since $Q_i(I, J) = 0$ whenever I or J is not a clique. So, to show that $D/8d \succeq Q_i \succeq -D/8d$, it is sufficient to show that for all vectors v with $v = \Pi v$ it happens that $|v^\top Q_i v| \leqslant v^T (D/8d) v$. To see this, let v_k be the part of v indexed by cliques of size exactly k. Now,

$$\begin{aligned}
|v^\top Q_i v| &\leqslant \sum_{k_1=0}^{d} \sum_{k_2=0}^{d} \|v_{k_1}\| \left\| \sum_{\substack{t \leqslant \tau \\ j \leqslant q}} M_j^{(k_1, k_2, t)} \right\| \|v_{k_2}\| \\
&\leqslant \sum_{k_1=0}^{d} \sum_{k_2=0}^{d} \frac{1}{100d^3} \left(2^{\binom{k_1}{2}/2+\binom{k_2}{2}/2} \|v_{k_1}\| \|v_{k_2}\| \right) \\
&\leqslant \sum_{k_1=0}^{d} \sum_{k_2=0}^{d} \frac{1}{200d^3} \left(2^{\binom{k_1}{2}} \|v_{k_1}\|^2 + 2^{\binom{k_2}{2}} \|v_{k_2}\|^2 \right) \\
&\leqslant \sum_{k=0}^{d} \frac{2^{\binom{k}{2}}}{100d^2} \|v_k\|^2 \leqslant v^\top (D/8d) v
\end{aligned}$$

□

5.6.3.1 Coefficient Decay in the Factorization: Proof of Lemma 5.6.10

We turn to the proof of Lemma 5.6.10, for which we want the following characterization of the effect of the separating factorization on the underlying graph of a ribbon.

We require the following combinatorial quantities:

Definitions for Lemma 5.6.12.

1. $I, J, S_\ell, S_r \subseteq [n]$ of size at most d.
2. Ribbons $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r$ satisfying 1,3*,2 but not 4 for $S_\ell, S_r, I, J \subseteq [n]$. (Remember that $\mathcal{R}_m$ may be improper.)
3. Ribbons $\mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r$ which are the separating factorization of $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r$, with separators S'_ℓ, S'_r. (Remember that $\mathcal{R}'_m$ may be improper.)
4. p, the number of vertex-disjoint paths from S_ℓ to S_r in $\mathcal{R}_m$.
5. p', the number of vertex-disjoint paths from S'_ℓ to S'_r in $\mathcal{R}'_m$.
6. $r = (|\mathcal{V}(\mathcal{R}_\ell)|+|\mathcal{V}(\mathcal{R}_m)|+|\mathcal{V}(\mathcal{R}_\ell)|-|S_\ell|-|S_r|)-(|\mathcal{V}(\mathcal{R}'_\ell)|+|\mathcal{V}(\mathcal{R}'_m)|+|\mathcal{V}(\mathcal{R}'_\ell)|-|S'_\ell|-|S'_r|)$, the number of intersections among R_ℓ, R_m, R_r.
7. $\mathfrak{D} = \mathcal{Z}(\mathcal{R}'_m) \setminus \mathcal{Z}(\mathcal{R}_m)$, the newly degree-0 (we write *isolated*) vertices in $\mathcal{R}'_m$.
8. $\mathfrak{U} \subseteq \mathcal{V}(\mathcal{R}_\ell) \cup \mathcal{V}(\mathcal{R}_m) \cup \mathcal{V}(\mathcal{R}_r)$, the set of vertices appearing in more than one of $\mathcal{V}(\mathcal{R}_\ell), \mathcal{V}(\mathcal{R}_m)$, and $\mathcal{V}(\mathcal{R}_r)$. Note that $\mathfrak{U} \subseteq \mathcal{V}(\mathcal{R}'_m)$.

Lemma 5.6.12.

$$\underbrace{|S'_\ell| + |S'_r| - (|S_\ell| + |S_r|)}_{\text{increase in separator size}} + \underbrace{p - p'}_{\text{lost paths between separators}} + \underbrace{|\mathfrak{D}|}_{\text{new isolated vertices}} \leqslant \underbrace{r}_{\text{number of intersections}} .$$

The following series of claims will help us in the proof of Lemma 5.6.12

Claim 5.6.13. $I \cap \mathcal{V}(\mathcal{R}'_m) \subseteq S'_\ell$ and $J \cap \mathcal{V}(\mathcal{R}'_m) \subseteq S'_r$.

Proof of claim. If $u \in I \cap \mathcal{V}(\mathcal{R}'_m)$ then since $I \subseteq \mathcal{V}(\mathcal{R}'_\ell)$, we have $u \in \mathcal{V}(\mathcal{R}'_\ell) \cap \mathcal{V}(\mathcal{R}'_m) = S'_\ell$, and similarly for the second part. □

Next we have a simple analysis of which vertices may possibly be newly isolated.

Claim 5.6.14. $\mathfrak{D} \subseteq \mathfrak{U}$.

Proof of claim. Let $u \in \mathfrak{D}$. If $u \in S_\ell$ or $u \in S_r$ we are done. Otherwise, if $u \in I$ or $u \in J$, then u appeared in more than one of $\mathcal{V}(\mathcal{R}_\ell), \mathcal{V}(\mathcal{R}_m), \mathcal{V}(\mathcal{R}_r)$ by the definition of the canonical factorization.

If neither of these cases hold, then u was incident to an edge in at least one of $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r$. Since that edge does not exist in $\mathcal{R}'_m$, it must have appeared at least twice among the edge sets of $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r$, and therefore u appeared at least twice among the vertex sets, thus proving the claim. □

Next we show that some vertices in $\mathfrak{U}$ cannot become isolated.

Claim 5.6.15. By Menger's theorem, there are $|S'_\ell|$ vertex-disjoint paths from $\mathfrak{U} \cap \mathcal{V}(\mathcal{R}_\ell)$ to I in $\mathcal{R}_\ell$. Let $u_\ell^{(1)}, \ldots, u_\ell^{(|S'_\ell|)}$ be distinct vertices so that $u^{(i)}$ is the last vertex in $\mathfrak{U}$ along the i-th vertex disjoint path. Let $u_r^{(1)}, \ldots, u_r^{(|S'_r|)}$ be similarly defined. None of the vertices u may be in $\mathfrak{D}$.

Proof of claim. Fix one of these vertices u, and consider its neighbor v one step farther along the path to I (or J). By definition, the vertex v does not appear in more than one of $\mathcal{V}(\mathcal{R}_\ell), \mathcal{V}(\mathcal{R}_m), \mathcal{V}(\mathcal{R}_r)$. If $v \in \mathcal{R}'_m$, then the edge (u, v) must be in $\mathcal{R}'_m$, and so u is not isolated in $\mathcal{R}'_m$. If $v \notin \mathcal{R}'_m$, then u must be in $S'_\ell \cup S'_r$, in which case by definition $u \notin \mathfrak{D}$. □

We set up sets q of vertices to divide up the intersecting vertices among $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r$ according to which ribbons witness the intersection.

Claim 5.6.16. Let

$$\begin{aligned}
q_{\ell,m,r} &\stackrel{\text{def}}{=} (\mathcal{V}(\mathcal{R}_r) \cap \mathcal{V}(\mathcal{R}_m) \cap \mathcal{V}(\mathcal{R}_\ell)) \setminus (S_\ell \cup S_r) \\
q_{\ell,r} &\stackrel{\text{def}}{=} (\mathcal{V}(\mathcal{R}_\ell) \cap \mathcal{V}(\mathcal{R}_r)) \setminus \mathcal{V}(\mathcal{R}_m) \\
q_{\ell,m} &\stackrel{\text{def}}{=} (\mathcal{V}(\mathcal{R}_\ell) \cap \mathcal{V}(\mathcal{R}_m)) \setminus (S_\ell \cup \mathcal{V}(\mathcal{R}_r)) \\
q_{r,m} &\stackrel{\text{def}}{=} (\mathcal{V}(\mathcal{R}_r) \cap \mathcal{V}(\mathcal{R}_m)) \setminus (S_r \cup \mathcal{V}(\mathcal{R}_\ell)) \,.
\end{aligned}$$

The sets q are pairwise disjoint, and

$$r = 2|q_{\ell,m,r}| + |q_{\ell,r}| + |q_{\ell,m}| + |q_{r,m}| + |S_\ell \cap (\mathcal{V}(\mathcal{R}_r) \setminus S_r)| + |S_r \cap (\mathcal{V}(\mathcal{R}_\ell) \setminus S_\ell)| \,.$$

Also, $\mathfrak{U} = q_{\ell,m,r} \cup q_{\ell,r} \cup q_{\ell,m} \cup q_{r,m} \cup S_\ell \cup S_r$.

Proof. By inspection. □

We are prepared to prove Lemma 5.6.12.

Proof of Lemma 5.6.12. We start by bounding the number of vertices in $\mathfrak{U} \setminus \mathfrak{D}$. By Claim 5.6.15, there are at least $|\{u_\ell^{(1)}, \ldots, u_\ell^{(|S'_\ell|)}, u_r^{(1)}, \ldots, u_r^{|S'_r|}\}|$ such vertices.

Let a be the number of pairs i, j so that $u_\ell^{(i)} = u_r^{(j)}$. Then there are vertex-disjoint paths $w_1, \ldots, w_a$ from S'_ℓ to S'_r. The path w corresponding to $u_\ell^{(i)} = u_r^{(j)}$ is given by following $u_\ell^{(i)}$'s path from I to $\mathcal{U}$, ending at $u_\ell^{(i)}$, then following $u_r^{(j)}$'s path from $\mathfrak{U}$ to J. This gives a path from I to J, which must have a subpath from S'_ℓ to S'_r.

Now consider the p vertex-disjoint paths from S_ℓ to S_r in $\mathcal{R}_m$. We claim that

$$p - |S_\ell \cap S_r| \leqslant |q_{\ell,m,r}| + |S_\ell \cap \mathcal{V}(\mathcal{R}_r) \setminus S_r| + |S_r \cap \mathcal{V}(\mathcal{R}_\ell) \setminus S_\ell| \\ + |\mathfrak{U} \setminus (\{u_\ell^{(1)}, \ldots, u_\ell^{(|S'_\ell|)}, u_r^{(1)}, \ldots, u_r^{|S'_r|}\} \cup \mathfrak{D})| + (p' - a) \tag{5.6.1}$$

In words, every nontrivial path from S_ℓ to S_r contributes to at least one of:

- $|q_{\ell,m,r}|$, the number of 3-way intersections,
- intersections between S_ℓ and $\mathcal{V}(\mathcal{R}_r)$ (but not S_r), intersections between $\mathcal{V}(\mathcal{R}_\ell)$ and S_r (but not S_ℓ),
- vertices in $\mathfrak{U}$ which are guaranteed not to become isolated (and which we have not yet accounted for), or
- vertex-disjoint paths from S'_ℓ to S'_r (which we have not yet accounted for).

Fix one such path. If it intersects $q_{\ell,m,r}$, $S_l \cap \mathcal{V}(\mathcal{R}_r)$, or $S_r \cap \mathcal{V}(\mathcal{R}_l)$ we are done, so suppose otherwise. If it is contained entirely in $q_{\ell,m} \cup q_{r,m} \cup (S_\ell \setminus \mathcal{V}(\mathcal{R}_r)) \cup (S_r \setminus \mathcal{V}(\mathcal{R}_l))$, then there is some edge along the path connecting a vertex in $\mathcal{V}(\mathcal{R}_\ell) \cap \mathcal{V}(\mathcal{R}_m) \setminus \mathcal{V}(\mathcal{R}_r)$ with one in $\mathcal{V}(\mathcal{R}_r) \cap \mathcal{V}(\mathcal{R}_m) \setminus \mathcal{V}(\mathcal{R}_\ell)$. That edge can occur nowhere else among $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r$, and so the incident vertices must not be in $\mathfrak{D}$. At the same time, if there is any vertex along the path which is outside $\mathfrak{U}$, then the nearest vertices along the path to either side which do lie in $\mathfrak{U}$ also must be outside $\mathfrak{D}$.

In either case, there are two vertices along the path in $\mathfrak{U} \setminus \mathfrak{D}$. If either of these is not among the u vertices, we are done. If both are, then by definition of the u vertices this creates a path from I to J, and so from S'_ℓ to S'_r. Furthermore, this path must be vertex

disjoint from the paths $w_1, \dots, w_a$ previously constructed, since the u vertices involved in those paths were $\mathcal{V}(\mathcal{R}_\ell) \cap \mathcal{V}(\mathcal{R}_r)$. This proves (5.6.1).

It's time to put things together. By Claim 5.6.14, we can bound $|\mathfrak{D}|$ by

$$|\mathfrak{D}| \leqslant |\mathfrak{U}| - |\mathfrak{U} \setminus \mathfrak{D}|.$$

We have $|\mathfrak{U} \setminus \mathfrak{D}| \geqslant |S'_\ell| + |S'_r| - a + |\mathfrak{U} \setminus (\{u_\ell^{(1)}, \dots, u_\ell^{(|S'_\ell|)}, u_r^{(1)}, \dots, u_r^{|S'_r|}\} \cup \mathfrak{D})|$, and $|\mathfrak{U}| = |q_{\ell,m,r}| + |q_{\ell,r}| + |q_{\ell,m}| + |q_{r,m}| + |S_\ell \cup S_r|$. This gives us

$$|\mathfrak{D}| \leqslant |q_{\ell,m,r}|+|q_{\ell,r}|+|q_{\ell,m}|+|q_{r,m}|+|S_\ell \cup S_r|-|S'_\ell|-|S'_r|+a-|\mathfrak{U}\setminus(\{u_\ell^{(1)}, \dots, u_\ell^{(|S'_\ell|)}, u_r^{(1)}, \dots, u_r^{|S'_r|}\}\cup\mathfrak{D})|\,.$$

Adding (5.6.1) to both sides and rearranging, we get

$$\begin{aligned} &p - p' + |\mathfrak{D}| \\ &\leqslant 2|q_{\ell,m,r}| + |S_\ell \cap (\mathcal{V}(\mathcal{R}_r) \setminus S_r)| + |S_r \cap (\mathcal{V}(\mathcal{R}_\ell) \setminus S_\ell)| \\ &+ |q_{\ell,r}| + |q_{\ell,m}| + |q_{r,m}| + |S_\ell \cup S_r| - |S'_\ell| - |S'_r| + |S_\ell \cap S_r|\,, \end{aligned}$$

and substituting $r = 2|q_{\ell,m,r}| + |S_\ell \cap (\mathcal{V}(\mathcal{R}_r) \setminus S_r)| + |S_r \cap (\mathcal{V}(\mathcal{R}_\ell) \setminus S_\ell)| + |q_{\ell,r}| + |q_{\ell,m}| + |q_{r,m}|$ gives

$$p - p' + |\mathfrak{D}| \leqslant r + |S_\ell \cup S_r| - |S'_\ell| - |S'_r| + |S_\ell \cap S_r|\,.$$

Notice that $|S_\ell \cup S_r| + |S_\ell \cap S_r| = |S_\ell| + |S_r|$, so we can rearrange to obtain the lemma. □

Now we can prove Lemma 5.6.10.

Proof of Lemma 5.6.10. First of all, we note that $c_i(\mathcal{R}_m)$ depends only on the shape of $\mathcal{R}_m$ by symmetry of our construction. We turn to the quantitative bound.

The proof is by induction. The coefficients $c_0(\mathcal{R}_m)$ are nonzero only for ribbons $\mathcal{R}_m$ which have $\mathcal{Z}(\mathcal{R}_m) = \emptyset$ and admitting $|S_\ell| = |S_r| = p$ paths from S_ℓ to S_r. Thus in the case that $i = 0$, the statement reduces to $c_0(\mathcal{R}_m) \leqslant 1$, which is true by definition.

Suppose the lemma holds for c_i, and consider c_{i+1}. By definition, for an (improper) S'_ℓ, S'_r-ribbon $\mathcal{R}'_m$ and ribbons $\mathcal{R}'_\ell, \mathcal{R}'_r$ satisfying 1 and 2,

$$c_{i+1}(\mathcal{R}'_m) = \sum_{\substack{\mathcal{R}_\ell,\mathcal{R}_m,\mathcal{R}_r \text{ satisfying} \\ 1,3^*,2 \text{ and not } 4 \text{ for some } S_\ell,S_r \\ r \text{ intersections outside } S_\ell,S_r \\ \text{separating factorization } \mathcal{R}'_\ell,\mathcal{R}'_m,\mathcal{R}'_r,S'_\ell,S'_r}} c_i(\mathcal{R}_m) \left(\frac{\omega}{n}\right)^r . \tag{5.6.2}$$

We introduce the shorthand $s' = \frac{|S'_\ell|+|S'_r|}{2}$. Consider first a particular term in the sum, $c_i(\mathcal{R}_m)(\omega/n)^r$, where $\mathcal{R}_m$ is an improper S_ℓ, S_r ribbon, and let $|\mathfrak{D}| = |\mathcal{Z}(\mathcal{R}'_m) \setminus \mathcal{Z}(\mathcal{R}_m)|$. By induction and Lemma 5.6.12,

$$\begin{aligned}
\left(\frac{\omega}{n}\right)^r \cdot c_i(\mathcal{R}_m) &\leqslant \left(\frac{\omega}{n}\right)^r \cdot \left(\frac{\omega}{n}\right)^s \cdot n^{\frac{p-|\mathcal{Z}(\mathcal{R}_m)|-i/2}{2}+\varepsilon s} && \text{by induction} \\
&= \left(\frac{\omega}{n}\right)^{s'} \cdot \left(\frac{\omega}{n}\right)^{r-s'+s} \cdot \cdot n^{\frac{p-|\mathcal{Z}(\mathcal{R}_m)|-i/2}{2}+\varepsilon s} \\
&= \left(\frac{\omega}{n}\right)^{s'} \cdot n^{-\varepsilon(r-s'+s)} \cdot n^{-\frac{1}{2}(r-s'+s)} \cdot n^{\frac{p-|\mathcal{Z}(\mathcal{R}_m)|-i/2}{2}+\varepsilon s} && \text{using } \omega = n^{1/2-\varepsilon} \\
&\leqslant \left(\frac{\omega}{n}\right)^{s'} \cdot n^{-\varepsilon(r-s'+s)} \cdot n^{-\frac{1}{2}(s'-s+p-p'+|\mathfrak{D}|)} \cdot n^{\frac{p-|\mathcal{Z}(\mathcal{R}_m)|-i/2}{2}+\varepsilon s} && \text{by Lemma 5.6.12} \\
&= \left(\frac{\omega}{n}\right)^{s'} \cdot n^{-\varepsilon(r-s'+s)} \cdot n^{\frac{p'-|\mathcal{Z}(\mathcal{R}'_m)|-i/2-s'+s}{2}+\varepsilon s} \\
&\text{canceling terms, using } |\mathcal{Z}(\mathcal{R}'_m)| = |\mathfrak{D}| + |\mathcal{Z}(\mathcal{R}_m)| \\
&= n^{-\varepsilon r} \cdot \left(\frac{\omega}{n}\right)^{s'} \cdot n^{\frac{p'-|\mathcal{Z}(\mathcal{R}'_m)|-i/2-(s'-s)}{2}+\varepsilon s'} \\
&\leqslant n^{-\varepsilon r} \cdot \left(\frac{\omega}{n}\right)^{s'} \cdot n^{\frac{p'-|\mathcal{Z}(\mathcal{R}'_m)|-(i+1)/2}{2}+\varepsilon s'} && \text{using } s'-s \geqslant 1/2, \text{ by Lemma 5.5.13}
\end{aligned}$$

Next we assess how many nonzero terms are in the sum (5.6.2) for a fixed r and a fixed $\mathcal{R}'_m$. For each vertex of $\mathcal{R}'_m$, there are 7 possibilities for which ribbon(s) it came from in $\{\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r\}$ so there are at most 7^τ choices overall (recall that $\mathcal{R}'_m$ has at most τ vertices for the terms we are looking at). Once we have chosen which ribbon(s) each vertex of $\mathcal{R}'_m$ came from, everything is fixed except for possible edges of $\mathcal{R}'_m$ which appear at least twice in $\mathcal{R}_\ell$, $\mathcal{R}_m$, and $\mathcal{R}_r$. There are two possibilities for each possible edge of $\mathcal{R}'_m$ which appears twice in $\mathcal{R}_\ell$, $\mathcal{R}_m$, and $\mathcal{R}_r$ and four possibilities for each possible edge of $\mathcal{R}'_m$ which appers three times in $\mathcal{R}_\ell$, $\mathcal{R}_m$, and $\mathcal{R}_r$. However, note that any such edge must be between an intersected vertex and either another intersected vertex or a vertex in $S_\ell \cup S_r$. Thus, there are at most $r\tau$ possible edges of $\mathcal{R}'_m$ which appear at least twice in $\mathcal{R}_\ell$, $\mathcal{R}_m$, and $\mathcal{R}_r$ and the total number of possibilities for these edges is at most $4^{r\tau}$.

All together there are at most $2^{O(r\tau)}$ nonzero terms for fixed r. This means that the total contribution from such terms is at most

$$2^{O(r\tau)} \cdot n^{-\varepsilon r} \cdot \left(\frac{\omega}{n}\right)^{s'} \cdot n^{\frac{p'-|\mathcal{Z}(\mathcal{R}'_m)|-(i+1)/2}{2}+\varepsilon s'}$$

As long as $\tau \leqslant (\varepsilon/C)\log n$ for some universal constant C, we have $2^{O(r\tau)} \cdot n^{-\varepsilon r} \ll 1/\tau$ for all $r \geqslant 1$. All in all, we obtain

$$c_{i+1}(\mathcal{R}'_m) \leqslant \left(\frac{\omega}{n}\right)^{s'} \cdot n^{\frac{p'-|\mathcal{Z}(\mathcal{R}'_m)|-(i+1)/2}{2}+\varepsilon s'}$$

which completes the induction. □

5.6.4 $\mathcal{L}\mathcal{L}^\top$ is Well-Conditioned — Proof of Lemma 5.6.3

In this section we prove Lemma 5.6.3, restated here.

Lemma (Restatement of Lemma 5.6.3)**.** *With high probability,* $\Pi\mathcal{L}\Pi\mathcal{L}^\top\Pi \succeq \Omega(\omega/n)^{d+1} \cdot \Pi$*, where as usual* Π *is the projector to* $\mathsf{Span}\{e_C : C \in C_{\leqslant d}\}$.

Proof of Lemma 5.6.3. We recall the definition of $\mathcal{L}$.

$$\mathcal{L}(I,S) = \left(\frac{\omega}{n}\right)^{-\frac{|S|}{2}} \sum_{\substack{\mathcal{R} \text{ having } 1 \\ |\mathcal{V}(\mathcal{R}_\ell)| \leqslant \tau}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_\ell)|} \chi_{\mathcal{R}_\ell}\,.$$

Consider a diagonal entry $\mathcal{L}(S,S)$. Since every ribbon $\mathcal{R}$ appearing in its expansion must have 1, in particular it has no edges inside S. Thus, by the same argument as in Lemma 5.4.4, with probability at least $1 - O(n^{-10\log n})$,

$$\mathcal{L}(S,S) = \left(\frac{\omega}{n}\right)^{\frac{|S|}{2}} (1 \pm n^{-\Omega(\varepsilon)})\,.$$

Let $\mathcal{L}^{\text{off-diag}}$ be given by

$$\mathcal{L}^{\text{off-diag}}(I,S) = \begin{cases} \mathcal{L}(I,S) \text{ if } I \neq S \\ 0 \text{ otherwise} \end{cases}\,.$$

We will consider the block of $\mathcal{L}^{\text{off-diag}}$ with rows indexed by sets of size s_ℓ and columns indexed by sets of size s_r for some $s_\ell, s_r \leqslant d$. For a fixed $t \leqslant \tau$, let $U_1^{(s_\ell,s_r,t)}, \ldots, U_q^{(s_\ell,s_r,t)}$ be all the graphs on vertex set $[t]$ with distinguished subsets of vertices A, B of size s_ℓ, s_r respectively, and where

- $A \neq B$,
- there are no edges inside B,
- every vertex in U outside $A \cup B$ is reachable from A without passing through B, and

- B is the unique minimum-size vertex separator in U separating A from B.

Then let $M_i^{(s_\ell,s_r,t)}$ be given by

$$M_i^{(s_\ell,s_r,t)}(I,S) = \left(\frac{\omega}{n}\right)^{t-\frac{s_r}{2}} \cdot \sum_{\mathcal{R} \text{ an } (I,S)\text{-ribbon with shape } U_i^{(s_\ell,s_r,t)}} \chi_{\mathcal{R}} \cdot$$

By assumption on $U_i^{(s_\ell,s_r,t)}$, there are s_r vertex-disjoint paths from A to B. Let $r = |A \cap B|$. By Lemma 5.6.8, with probability at least $1 - O(n^{-100\log n})$,

$$\begin{aligned}\left\|M_i^{(s_\ell,s_r,t)}\right\| &\leqslant \left(\frac{\omega}{n}\right)^{\frac{s_r}{2}} \cdot \left(\frac{\omega}{n}\right)^{t-s_r} \cdot n^{\frac{t-s_r}{2}} \cdot 2^{O(t)} \cdot (\log n)^{O(t-r+(s_r-r))} \\ &= \left(\frac{\omega}{n}\right)^{\frac{s_r}{2}} \cdot n^{-\varepsilon(t-s_r)} \cdot 2^{O(t)} \cdot (\log n)^{O(t-s_r)},\end{aligned}$$

where in the last step we have used that $t \geqslant s_\ell + s_r - r$ and $s_r \leqslant s_\ell$, which holds by the vertex-separator requirement on B. There are at most $2^{\binom{t}{2}-\binom{s_r}{2}+O(t)}$ choices for $U_i^{(s_\ell,s_r,t)}$ when s_ℓ, s_r, t are fixed, by the requirement that U have no edges inside B. Summing over all q for a fixed t, we get by triangle inequality

$$\left\|\sum_{i\leqslant q} M_i^{(s_\ell,s_r,t)}\right\| \leqslant \left(\frac{\omega}{n}\right)^{\frac{s_r}{2}} \cdot 2^{\binom{t}{2}-\binom{s_r}{2}+O(t)} \cdot n^{-\varepsilon(t-s_r)} \cdot (\log n)^{O(t-s_r)}$$

with probability $1 - O(n^{-99\log n})$. By our assumptions on d, τ, and ε, this is at most $(\omega/n)^{s_r/2} \cdot 1/d^4$.

The following standard manipulations now prove the lemma. Let $D' \in \mathbb{R}^{\binom{[n]}{\leqslant d}}$ be the diagonal matrix with $D'(S,S) = (\omega/n)^{|S|/2}$ if S is a clique in G and 0 otherwise. Then we can decompose $\mathcal{L} = D + E + \mathcal{L}^{\text{off-diag}}$, where E is a diagonal matrix with $|E(S,S)| \leqslant n^{-\Omega(\varepsilon)} \cdot (\omega/n)^{|S|/2}$. Then we have

$$\Pi\mathcal{L}\Pi\mathcal{L}^\top\Pi = D^2$$

$$+ \Pi(D\Pi\mathcal{L}^{\text{off-diag}} + D\Pi E + E\Pi D + E\Pi\mathcal{L}^{\text{off-diag}} + \mathcal{L}^{\text{off-diag}}\Pi D + \mathcal{L}^{\text{off-diag}}\Pi E$$
$$+ E\Pi E + \mathcal{L}^{\text{off-diag}}\Pi\mathcal{L}^{\text{off-diag}})\Pi$$

Each of the above matrices aside from D^2 is a $d \times d$ block matrix, where the (s_ℓ, s_r) block is $\binom{[n]}{s_\ell} \times \binom{[n]}{s_r}$ dimensional and has norm at most $(\omega/n)^{(s_\ell+s_r)/2} \cdot d^{-4}$. By the same argument as in the proof of Lemma 5.6.2, using Cauchy-Schwarz to combine the d^2 blocks, we obtain the lemma. □

5.6.5 High-Degree Matrices Have Small Norms

In this section we prove Lemma 5.6.4, restated here:

Lemma (Restatement of Lemma 5.6.4)**.** *With high probability,* $\|\xi_0 - \ldots + \xi_{2d}\| \leqslant n^{-16d}$.

We recall the definition of ξ_i. For a coefficient function on ribbons $c_{i-1}(\mathcal{R}_m)$, we have a matrix $\mathcal{E}$ given by

$$\mathcal{E}(I,J) = \sum_{\substack{S_\ell, S_r \subseteq [n] \\ |S_\ell|, |S_r| \leqslant d}} \left(\frac{\omega}{n}\right)^{-\frac{|S_\ell|+|S_r|}{2}}$$
$$\sum_{\substack{\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r \text{ satisfying} \\ 1,3^*,2 \text{ and not } 4 \\ |\mathcal{V}(\mathcal{R}_\ell)|, |\mathcal{V}(\mathcal{R}_m)|, |\mathcal{V}(\mathcal{R}_r)| \leqslant \tau \\ \text{separating factorization} \\ \mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r, S'_\ell, S'_r}} c_{i-1}(\mathcal{R}_m) \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_\ell)|+|\mathcal{V}(\mathcal{R}_r)|+|\mathcal{V}(\mathcal{R}_m)|-\frac{|S_\ell|+|S_r|}{2}} \cdot \chi_{\mathcal{R}'_\ell} \cdot \chi_{\mathcal{R}'_m} \cdot \chi_{\mathcal{R}'_r},$$

and another one, $\mathcal{E}'$, given by

$$\mathcal{E}'(I,J) = \sum_{\substack{S_\ell, S_r \subseteq [n] \\ |S_\ell|, |S_r| \leqslant d}} \left(\frac{\omega}{n}\right)^{-\frac{|S_\ell|+|S_r|}{2}}$$

$$\sum_{\substack{\mathcal{R}_\ell,\mathcal{R}_m,\mathcal{R}_r \text{ satisfying} \\ 1,3^*,2 \text{ and not } 4 \\ \text{separating factorization} \\ \mathcal{R}'_\ell,\mathcal{R}'_m,\mathcal{R}'_r,S'_\ell,S'_r \\ |\mathcal{V}(\mathcal{R}'_\ell)|,|\mathcal{V}(\mathcal{R}'_m)|,|\mathcal{V}(\mathcal{R}'_r)|\leqslant\tau}} c_{i-1}(\mathcal{R}_m)\left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_\ell)|+|\mathcal{V}(\mathcal{R}_r)|+|\mathcal{V}(\mathcal{R}_m)|-\frac{|S_\ell|+|S_r|}{2}} \cdot \chi_{\mathcal{R}'_\ell} \cdot \chi_{\mathcal{R}'_m} \cdot \chi_{\mathcal{R}'_r} \cdot$$

Then the matrix ξ_i is given by $\mathcal{E} - \mathcal{E}'$.

We will actually prove a bound on the Frobenious norm of each matrix ξ_i. The following will allow us to control the magnitude of the entries. It follows immediately from our concentration bound Lemma A.0.1, which is proved via the moment method. (Under the slightly stronger assumption $\tau \ll \varepsilon \log n / \log\log n$, it would also follow from standard hypercontractivity.)

Lemma 5.6.17. *Suppose c_T are a collection of coefficients, one for each $T \subseteq \binom{[n]}{2}$, and there is a constant C such that*

1. *If $|T| > C\tau$ then $c_T = 0$.*
2. *Otherwise, $|c_T| \leqslant (\omega/n)^{|T|/C - Cd}$.*

Then with probability at least $1 - O(n^{-100\log n})$ it occurs that $\left|\sum_{T\subseteq\binom{[n]}{2}} c_T \cdot \chi_T\right| \leqslant n^{-20d}$.

We will also need several facts about the coefficients of ribbons in the expansion of each matrix ξ_i.

Lemma 5.6.18. *Every triple $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r$ appearing with nonzero coefficient in ξ_c satisfies $|\mathcal{V}(\mathcal{R}_\ell)| + |\mathcal{V}(\mathcal{R}_m)| + |\mathcal{V}(\mathcal{R}_r)| = \Theta(\tau)$.*

Proof. To appear with nonzero coefficient, the triple $\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r$ with separating factorization $\mathcal{R}'_\ell, \mathcal{R}'_m, \mathcal{R}'_r$ must either have

$$|\mathcal{V}(\mathcal{R}_\ell)|, |\mathcal{V}(\mathcal{R}_m)|, |\mathcal{V}(\mathcal{R}_r)| \leqslant \tau \quad \text{but} \quad |\mathcal{V}(\mathcal{R}'_\ell)| > \tau \text{ or } |\mathcal{V}(\mathcal{R}'_m)| > \tau \text{ or } |\mathcal{V}(\mathcal{R}'_\ell)| > \tau\,,$$

or

$$|\mathcal{V}(\mathcal{R}'_\ell)|, |\mathcal{V}(\mathcal{R}'_m)|, |\mathcal{V}(\mathcal{R}'_r)| \leqslant \tau \quad \text{but} \quad |\mathcal{V}(\mathcal{R}_\ell)| > \tau \text{ or } |\mathcal{V}(\mathcal{R}_m)| > \tau \text{ or } |\mathcal{V}(\mathcal{R}_\ell)| > \tau\,.$$

In the first case, we must have one of $|\mathcal{V}(\mathcal{R}_\ell)| \geqslant \tau/3$ or $|\mathcal{V}(\mathcal{R}_m)| \geqslant \tau/3$ or $|\mathcal{V}(\mathcal{R}_r)| \geqslant \tau/3$. In the second, we must have $|\mathcal{V}(\mathcal{R}_\ell)|, |\mathcal{V}(\mathcal{R}_m)|, \mathcal{V}(\mathcal{R}_r)| \leqslant 3\tau$. □

We are prepared to prove Lemma 5.6.4.

Proof of Lemma 5.6.4. We will apply Lemma 5.6.17 to $\xi_i(I, J)$ for each $i \leqslant 2d$ and $I, J \subseteq [n]$ with $|I|, |J| \leqslant d$. So consider the Fourier expansion of $\xi_i(I, J)$, given by

$$\xi_i(I, J) = \sum_{T \subseteq \binom{[n]}{2}} c_T \cdot \chi_T\,.$$

From Lemma 5.6.18, we obtain that if $|T| > C\tau$ then $c_T = 0$, for some absolute constant C. For smaller T we need a bound on the magnitude $|c_T|$. The coefficient c_T is bounded by

$$|c_T| \leqslant \sum_{\substack{S_\ell, S_r \subseteq [n] \\ |S_\ell|, |S_r| \leqslant d}} \left(\frac{\omega}{n}\right)^{-\frac{|S_\ell|+|S_r|}{2}} \sum_{\substack{\mathcal{R}_\ell, \mathcal{R}_m, \mathcal{R}_r \\ \text{nonzero in } \xi_i(I,J) \text{ as in 5.6.18} \\ \chi_{\mathcal{R}_\ell} \cdot \chi_{\mathcal{R}_m} \cdot \chi_{\mathcal{R}_r} = \chi_T}} c_{i-1}(\mathcal{R}_m) \left(\frac{\omega}{n}\right)^{|\mathcal{V}(\mathcal{R}_\ell)|+|\mathcal{V}(\mathcal{R}_r)|+|\mathcal{V}(\mathcal{R}_m)|-\frac{|S_\ell|+|S_r|}{2}} \tag{5.6.3}$$

By Lemma 5.6.10, we have $c_{i-1}(\mathcal{R}_m) \leqslant n^d \leqslant (\omega/n)^{-2d}$. At the same time, there are at most $2^{O(\tau^2)}$ nonzero terms in the sum (5.6.3). Thus by Lemma 5.6.18 and our assumptions on d, τ, and ε, the coefficient c_T is at most $(\omega/n)^{\tau/C - Cd}$ for some absolute constant C.

Applying Lemma 5.6.17, we obtain $|\xi_i(I,J)| \leqslant n^{-20d}$ with probability $1 - O(n^{-100\log n})$. Taking a union bound over all $n^{2d} \leqslant n^{2\log n}$ entries of ξ_i, and over all $i \leqslant 2d$, we obtain that $\|\xi_0 - \ldots + \xi_{2d}\| \leqslant \|\xi_0 - \ldots + \xi_{2d}\|_F \leqslant n^{-16d}$ with probability $1 - O(n^{-96\log n})$. □

Appendices

Appendix A

Omitted Details from Chapter 5

A.0.1 Calibration of $\tilde{\mathbb{E}}$

In this subsection we prove Lemma 5.4.3, restated here.

Lemma (Restatement of Lemma 5.4.3)**.** *Let $f_G(x) = \sum_{|S|\leqslant 2d} c_S(G) \cdot x_S$ be a real-valued polynomial on $\{0,1\}^n$ whose coefficients have degree at most τ when expressed in the ± 1 indicators G_e for edges in G. Then, $\mathbb{E}_{G\sim G(n,\frac{1}{2})}[\tilde{\mathbb{E}}[f_G(x)]] = \mathbb{E}_{(H,x)\sim G(n,1/2,\omega)}[f_H(x)]$.*

Proof. The proof is straightforward by expanding the coefficients f in the Fourier basis. For $S \subseteq [n]$, let $c_S : G \mapsto \mathbb{R}$ be maps so that $f_G(x) = \sum_{S\subseteq[n]} c_S \cdot x_S$.

$$\begin{aligned}
\mathop{\mathbb{E}}_{G\sim G(n,\frac{1}{2})}[\tilde{\mathbb{E}}[f_G(x)]] &= \mathop{\mathbb{E}}_{G\sim G(n,\frac{1}{2})}\left[\tilde{\mathbb{E}}\left[\sum_{S\subseteq[n]} c_S \cdot x_S\right]\right] \\
&= \sum_{S\subseteq[n]} \mathop{\mathbb{E}}_{G\sim G(n,\frac{1}{2})}\left[c_S\, \tilde{\mathbb{E}}[x_S]\right] \\
&= \sum_{S\subseteq[n]} \mathop{\mathbb{E}}_{G\sim G(n,\frac{1}{2})}\left[\sum_{T,T'\subseteq\binom{[n]}{2}} \widehat{c_S}(T)\, \widehat{\tilde{\mathbb{E}}}(T') \cdot \chi_T\chi_{T'}\right] \\
&= \sum_{S\subseteq[n]} \sum_T \widehat{c_S}(T) \mathop{\mathbb{E}}_{(H,x)\sim G(n,1/2,\omega)}[\chi_T(H)\cdot x_S] \\
&= \mathop{\mathbb{E}}_{(H,x)\sim G(n,1/2,\omega)}\left[\sum_{S\subseteq[n]} \sum_T \widehat{c_S}(T)\chi_T(H)\prod_{i\in S} x_i\right]
\end{aligned}$$

$$= \mathop{\mathbb{E}}_{(H,x)\sim G(n,1/2,\omega)} \left[\sum_{S\subseteq[n]} c_S \prod_{i\in S} x_i \right]$$
$$= \mathop{\mathbb{E}}_{(H,x)\sim G(n,1/2,\omega)} [f_H(x)] . \qquad \square$$

A.0.2 Concentration Bounds for Linear Constraints

In this section we prove Lemma 5.4.4. We will use the following elementary concentration bound repeatedly. (It is the scalar version of the matrix concentration bound Lemma 5.6.8; we state and prove a scalar version here because it is a good warmup for Lemma 5.6.8.)

Lemma A.0.1. *Let $\mathcal{T}$ be a family of subsets of $\binom{[n]}{2}$ so that for every $T, T' \in \mathcal{T}$ there exists $\sigma : [n] \to [n]$ a permutation of vertices so that $\sigma(T) = T'$. Let t be the number of vertices incident to edges in any $T \in \mathcal{T}$. For every $s \geqslant 0$ and every even ℓ,*

$$\mathop{\mathbb{P}}_{G\sim G(n,1/2)} \left\{ \left| \sum_{T\in\mathcal{T}} \chi_T(G) \right| \leqslant s \right\} \geqslant 1 - \frac{n^{t\ell/2} \cdot (t\ell)^{t\ell}}{s^\ell} .$$

Proof. Let $\ell \in \mathbb{N}$ be a parameter to be chosen later. We will estimate $\mathbb{E}_{G\sim G(n,1/2)}[(\sum_{T\in\mathcal{T}} \chi_T)^\ell]$.

$$\mathop{\mathbb{E}}_{G\sim G(n,1/2)} \left[\left(\sum_{T\in\mathcal{T}} \chi_T \right)^\ell \right] = \sum_{T_1,\ldots,T_\ell\in\mathcal{T}} \mathop{\mathbb{E}}_{G\sim G(n,1/2)} \prod_{j\leqslant\ell} \chi_{T_j}$$
$$= |\{(T_1,\ldots,T_\ell) : \mathbb{E} \prod_{j\leqslant\ell} \chi_{T_j} = 1\}| .$$

In order to have $\mathbb{E}\prod_{j\leqslant\ell} \chi_{T_j} = 1$, every edge in the multiset $\bigcup_{j\leqslant\ell} T_j$ must appear at least twice, so every vertex in the multiset $\bigcup_{j\leqslant\ell} \mathcal{V}(T_j)$ also appears at least twice. Thus, this multiset contains at most $t\ell/2$ distinct vertices. Since each $T_j \in \mathcal{T}$, each is uniquely

determined by an ordered tuple of t elements of $[n]$. Thus, there are at most $n^{t\ell/2} \cdot (t\ell)^{t\ell}$ distinct choices for $(T_1, \dots, T_\ell)$, so

$$\mathop{\mathbb{E}}_{G\sim G(n,1/2)} \left[\left(\sum_{T\in\mathcal{T}} \chi_T \right)^\ell \right] \leqslant n^{t\ell/2} \cdot (t\ell)^{t\ell}.$$

For even ℓ, by Markov's inequality,

$$\begin{aligned} \mathbb{P}\left\{ \left| \sum_{T\in\mathcal{T}} \chi_T \right| > s \right\} &= \mathbb{P}\left\{ \left| \sum_{T\in\mathcal{T}} \chi_T) \right|^\ell > s^\ell \right\} \\ &\leqslant \frac{n^{t\ell/2} \cdot (t\ell)^{t\ell}}{s^\ell}. \end{aligned}$$

□

Lemma (Restatement of Lemma 5.4.4)**.** *With high probability,* $\tilde{\mathbb{E}}[1] = 1 \pm n^{-\Omega(\varepsilon)}$ *and* $\tilde{\mathbb{E}}[\sum_{i\in[n]} x_i] = \omega \cdot (1 \pm n^{-\Omega(\varepsilon)})$.

Proof. We will prove the statement regarding $\tilde{\mathbb{E}}[1]$; the bound for $\tilde{\mathbb{E}}[\sum_{i\in[n]} x_i]$ is almost identical.

Recall the Fourier expansion

$$\tilde{\mathbb{E}}[1] - 1 = \sum_{\substack{T\subseteq\binom{[n]}{2} \\ 2\leqslant|\mathcal{V}(T)|\leqslant\tau}} \left(\frac{\omega}{n}\right)^{|\mathcal{V}(T)|} \cdot \chi_T \,.$$

Considering each $T \subseteq \binom{[n]}{2}$ as a graph, we partition $\{T \subseteq \binom{[n]}{2} : |\mathcal{V}(T)| = t\}$ into p_t families $\{\mathcal{T}_i^t\}_{i=1}^p$ by placing T and T' in the same family iff there exists a permutation $\sigma : [n] \to [n]$ of vertices so that $\sigma(T) = T'$. Thus,

$$\tilde{\mathbb{E}}[1] - 1 = \sum_{t=2}^{\tau} \left(\frac{\omega}{n}\right)^t \sum_{i=1}^{p_t} \sum_{T\in\mathcal{T}_i^t} \chi_T \leqslant \sum_{t=2}^{\tau} \left(\frac{\omega}{n}\right)^t \sum_{i=1}^{p_t} \left| \sum_{T\in\mathcal{T}_i^t} \chi_T \right| .$$

By Lemma A.0.1 (taking $\ell = (\log n)^2$), and since $t \leqslant \tau \leqslant \log n$, each $\mathcal{T}_i^t$ satisfies

$$\mathbb{P}\left\{\left|\sum_{T\in\mathcal{T}_i^t} \chi_T\right| < O(n^{t/2}\cdot(\log n)^{3t})\right\} \geqslant 1-(\tau\cdot 2^{t^2}\cdot n^{\log n})^{-1}\,.$$

By a union bound over all $p_t \leqslant 2^{t^2}$ families $\mathcal{T}_i^t$, we get that with high probability,

$$|\,\tilde{\mathbb{E}}[1]-1| \leqslant \tau\cdot\max_{t\leqslant\tau}\left(2^{t^2}\cdot\left(\frac{\omega}{\sqrt{n}}\right)^t\right).$$

For $\tau \leqslant (\varepsilon/2)\log n$ and $\omega = n^{1/2-\varepsilon}$, this is at most $n^{-\Omega(\varepsilon)}$. □

.

A.0.3 Combinatorial Proofs about Ribbons

In this section we prove Lemma 5.5.3, restated here:

Lemma (Restatement of Lemma 5.5.3)**.** *Let $\mathcal{R}$ be an (I,J)-ribbon. There is a unique minimum vertex separator S of $\mathcal{R}$ such that S separates I and Q for any vertex separator Q of $\mathcal{R}$. We call S the* leftmost *separator in $\mathcal{R}$. We define the* rightmost *separator analogously and we denote them by $S_L(\mathcal{R})$ and $S_R(\mathcal{R})$ respectively.*

We start by defining a natural partial order on the set of vertex separators in a ribbon $\mathcal{R}$.

Definition A.0.2. We write $Q_1 \leqslant Q_2$ for two vertex separators Q_1 and Q_2 of an (I,J)-ribbon $\mathcal{R}$ if Q_1 separates I and Q_2.

Next, we check that the definition above indeed is a partial order.

Lemma A.0.3. *For any set of minimum vertex separators Q_1, Q_2, Q_3 an (I, J)-ribbon, we have:*

1. $Q_1 \preccurlyeq Q_1$.
2. *If* $Q_1 \preccurlyeq Q_2$ *and* $Q_2 \preccurlyeq Q_3$, *then,* $Q_1 \preccurlyeq Q_3$.
3. *If* $Q_1 \preccurlyeq Q_2$ *and* $Q_2 \preccurlyeq Q_1$, *then,* $Q_1 = Q_2$.

Proof. The first statement is immediate from the definition. For the second, consider a path P from I to Q_3 in $\mathcal{R}$. Since $Q_2 \preccurlyeq Q_3$, P passes through a vertex in Q_2. Thus, P contains a subpath that connects I and Q_2. But since $Q_1 \preccurlyeq Q_2$, this subpath must pass through Q_1. Thus, any such P must pass through Q_1 and thus, $Q_1 \preccurlyeq Q_3$.

Finally, for the third statement, let $k = |Q_1| = |Q_2|$. Then, using Menger's theorem (Fact 5.3.2, there is a set of k vertex disjoint paths $P_1, P_2, \ldots, P_k$ between I and J. By virtue of Q_1, Q_2 being *minimum* vertex separators of $\mathcal{R}$, Q_1 and Q_2 must intersect each P_i in exactly one vertex. It is then immediate that the only way $Q_1 \preccurlyeq Q_2$ and $Q_2 \preccurlyeq Q_1$ if every P_i intersects Q_1, Q_2 in the same vertex. □

Now we can prove Lemma 5.5.3.

Proof of Lemma 5.5.3. It is enough to show that for any two minimum separators Q_1, Q_2 of size k in R, there are separators Q_L, Q_R such that $Q_L \preccurlyeq Q_1 \preccurlyeq Q_R$ and $Q_L \preccurlyeq Q_2 \preccurlyeq Q_R$. We now construct Q_L and Q_R as required.

Let $U = Q_1 \cap Q_2$ and $V = Q_1 \Delta Q_2$. Let $W_L \subseteq V$ be the set of vertices w such that there is a path from I to w that doesn't pass through $Q_1 \cup Q_2$. Similarly, let $W_R \subseteq V$ be

the set of vertices such that there is a path from w to some vertex in J that doesn't pass through any vertex in $Q_1 \cup Q_2$. Then we first observe:

Claim A.0.4. $W_L \cap W_R = \emptyset$.

Proof of Claim. Assume otherwise and let $w \in W_L \cap W_R$. Then there is a path between I and J that doesn't go through any vertex in at least one of Q_1 or Q_2 contradicting that both are in fact vertex separators. □

Next, we have:

Claim A.0.5. Let $Q_L = U \cup W_L$ and $Q_R = U \cup W_R$. Then Q_L, Q_R are both vertex separators in R.

Proof of Claim. We only give the argument for Q_L, the other case is similar. Assume there is a path P from I to J that does not pass through Q_L. P must intersect $Q_1 \cup Q_2$. Then there is a vertex $v \in Q_1 \cup Q_2$ such that there is a path I to v which intersects no other vertices in $Q_1 \cup Q_2$. This implies that either $v \in U$ or $v \in W_L$. But by our construction of W_L this is a contradiction. □

Finally, we note that both Q_L, Q_R must in fact be *minimum* vertex separators.

Claim A.0.6. $|Q_L| = |Q_R| = |Q_1| = |Q_2| = k$

Proof of Claim. Let $|Q_1| = |Q_2| = k$. Then $2k = |Q_1| + |Q_2| = 2|U| + |V| \geqslant 2|U| + |W_L| + |W_R| = |U \cup W_L| + |U \cup W_R| = |Q_L| + |Q_R|$. Since Q_L and Q_R are vertex separators, $|Q_L|, |Q_R| \geqslant k$. Thus, $|Q_L| = |Q_R| = k$. □

Finally, we have the ordering requirement on Q_L and Q_R.

Claim A.0.7. $Q_L \preccurlyeq Q_1$ and $Q_2 \preccurlyeq Q_R$.

Proof of Claim. Let P be a path from I to Q_1, let v be the first vertex on this path which is in $Q_1 \cup Q_2$. Then, $v \in U$ or $v \in W_L$. Thus, $Q_L \preccurlyeq Q_1$. The other case is similar. □

This concludes the proof of the lemma. □

A.1 Spectral Norms

The results in this section are in essence due to Medarametla and Potechin [84]. For completeness, we state and prove them here in the language and notation of the current paper, with minor modifications as needed.

Lemma (Restatement of Lemma 5.6.8)**.** *Let U be a graph on $t \leqslant O(\log n)$ vertices, with two distinguished subsets of vertices A and B, and suppose:*

- *U admits p vertex-disjoint paths from $A \setminus B$ to $B \setminus A$.*
- *$|A \cap B| = r$.*
- *Every vertex outside $A \cup B$ has degree at least 1.*

Let $M = M(G)$ be the graphical matrix with shape U. Then, whp, $\|M\| \leqslant n^{\frac{t-p-r}{2}} \cdot 2^{O(t)} \cdot (\log n)^{O(t-r+p)}$.

Proof of Lemma 5.6.8. We proceed by the trace power method, with a dependence-breaking step beforehand.

Breaking Dependence. Let $q_1, \ldots, q_p$ be vertex-disjoint paths from $A \setminus B$ to $B \setminus A$ in U. Without loss of generality we can take each to intersect $A \setminus B$ and $B \setminus A$ only at its endpoints. We will partition the space of labelings σ into disjoint sets $S_1, \ldots, S_m$. For each S_k there will be a partition V_1^k, V_2^k of $[n]$ so that $\sigma(\bigcup_{j \leqslant p} q_j) \subseteq V_1^k$ and $\sigma(U \setminus (\bigcup_{j \leqslant p} q_j)) \subseteq V_2^k$ for every $\sigma \in S_k$. Let $(V_1^1, V_2^1), \ldots, (V_1^m, V_2^m)$ be a sequence of independent uniformly random partitions of $[n]$. Call a labeling σ *good* at k if the preceeding conditions apply to σ for the partition V_1^k, V_2^k and not for any $V_1^{k'}, V_2^{k'}$ for some $k' < k$. Let $S_k = \{\sigma : \sigma \text{ is good at } k\}$.

Claim A.1.1. There is $m = O(2^t \cdot t \cdot \log n)$ so that $\bigcup_{k=1}^m S_k$ contains every labeling $\sigma : U \to G$.

Proof. For a fixed σ,

$$\mathbb{P}\{\sigma \text{ not good for some } k \leqslant m\} \leqslant (1 - 2^{-t})^m$$

since every vertex $u \in U$ is in V_i with probability $1/2$. If $m \geqslant 10t2^t \log n$, then by a union bound over all $\sigma : U \to G$ (of which there are at most n^t), we get $\mathbb{P}\{$ all σ good for some $k \leqslant m \} > 0$. □

Henceforth, let $S_1, \ldots, S_m$ be the partition guaranteed by the preceeding claim. For $k \leqslant m$, let $M_k(I, J) = \sum_{\sigma \in S_k : \sigma(A)=I, \sigma(B)=J} \mathrm{val}(\sigma)$. Then $M = \sum_{k=1}^m M_k$.

Moment Calculation. Let $\ell = \ell(n)$ be a parameter to be chosen later. By the triangle inequality, $\|M\| \leqslant \sum_{k=1}^m \|M_k\|$. Fix k. We expand $\mathbb{E}_G \operatorname{Tr}(M_k^\top M_k)^\ell$ as

$$\mathbb{E} \operatorname{Tr}(M_k^\top M_k)^\ell = \mathbb{E} \sum_{\substack{\sigma_1, \ldots, \sigma_{2\ell} \in S_k \\ \sigma_{2i}(A) = \sigma_{2i-1}(A) \\ \sigma_{2i}(B) = \sigma_{2i+1}(B)}} \prod_{j=1}^{2\ell} \mathrm{val}(\sigma_j)\,.$$

(Here arithmetic with indices i is modulo 2ℓ, so for example we take $2i+1 = 1$.) For any σ,

$$\mathrm{val}(\sigma) = \prod_{(i,j)\in U} G_{\sigma(i),\sigma(j)}\,.$$

Notice that for all $\sigma_1, \ldots, \sigma_{2\ell}$, the expectation $\mathbb{E}\prod_{j=1}^{2\ell} \mathrm{val}(\sigma_j)$ is either 0 or 1. We will bound the number of $\sigma_1, \ldots, \sigma_{2\ell}$ for which $\mathbb{E}\prod_{j=1}^{2\ell} \mathrm{val}(\sigma_j) = 1$ by bounding the number of distinct labels such a family of labelings may assign to vertices in U.

Fix $\sigma_1, \ldots, \sigma_{2\ell} \in S_k$. Consider the family $q_1, \ldots, q_p$ of vertex-disjoint paths. Every edge in every q_j receives one pair of labels from each σ_i. Consider these labels arranged on 2ℓ adjoined copies of each q_j, one for each σ (giving p paths with $2\ell \sum_{j\leqslant p} |q_j|$ edges in total, where $|q_j|$ is the number of edges in q_j). Every pair of labels $\{\sigma_i(v), \sigma_i(w)\}$ appearing on an edge (v, w) in this graph must also appear on some distinct edge (v', w') in order to have $\mathbb{E}\prod_{i=1}^{2\ell} \mathrm{val}(\sigma_i) = 1$; otherwise the disjointness of V_1^k, V_2^k would be violated. Merging edges which received the same pair of labels, we arrive at a graph with at most p connected components and at most $\ell \sum_{j\leqslant p} |q_j|$ edges, and so at most $\ell \sum_{j\leqslant p} |q_j| + p$ vertices. Thus, the vertices in $q_1, \ldots, q_p$ together receive at most $\ell \sum_{j\leqslant p} |q_j| + p$ distinct labels among all $\sigma_1, \ldots, \sigma_{2\ell}$.

Next we account for labels of $v \notin (\bigcup_{j\leqslant p} q_j \cup A \cup B)$. If $\mathbb{E}_G \prod_{i=1}^{2\ell} \mathrm{val}(\sigma_i) = 1$ then the 2ℓ-size multiset $\{\sigma_i(v)\}_{i\leqslant 2\ell}$ of labels for such v contains at most ℓ distinct labels, since by assumption v has degree at least 1 in U.

Next we account for labels of vertices in $A \setminus (B \cup \bigcup_{j\leqslant p} q_j)$ and $B \setminus (A \cup \bigcup_{j\leqslant p} q_j)$. Every such vertex receives a label from every σ_i, but σ_{2i} and σ_{2i-1} must agree on A-labels and σ_{2i} and σ_{2i+1} must agree on B-labels. So in total there are at most $\ell(|A| + |B| - 2p - 2r)$ distinct labels for such vertices.

This means that among the labels $\sigma_i(j)$ for all $j \notin A \cap B$, there are at most

$$\underbrace{\ell \sum_{j \leqslant p} |q_j| + p}_{\text{labels from paths}} + \underbrace{\ell(|A| + |B| - 2p - 2r)}_{\text{additional vertices in } A \cup B \setminus (A \cap B)} + \underbrace{\ell(t - (|A| + |B| - r) - (\sum_j |q_j| - p))}_{\text{vertices in } U \setminus (\bigcup_j q_j \cup A \cup B)} = \ell(t-p-r)+p$$

unique labels.

Finally, consider the labels of the r vertices $j_1, \ldots, j_r$ in $A \cap B$. The first labelling σ_1 assigns these vertices some $\sigma_1(j_1), \ldots, \sigma_1(j_r)$ labels in G. Since σ_2 agrees with σ_1 on A-vertices, we must have $\sigma_2(j_1) = \sigma_1(j_1), \ldots, \sigma_1(j_r) = \sigma_2(j_r)$. Since σ_3 agrees with σ_2 on B-vertices, we must have $\sigma_3(j_1) = \sigma_2(j_1), \ldots, \sigma_3(j_r) = \sigma_2(j_r)$, and so on. So there are at most r unique labels for such vertices.

Now we can assess how many choices there are for $\sigma_1, \ldots, \sigma_{2\ell} \in S_k$ so that $\mathbb{E} \prod_{i \leqslant 2\ell} \mathrm{val}(\sigma_i) = 1$. To choose such a collection $\sigma_1, \ldots, \sigma_{2\ell}$, we proceed in stages.

Stage 1. Choose the labels $\sigma_i(j_1), \ldots, \sigma_i(j_r)$ of all the vertices in $A \cap B$. Here there are at most n^r options.

Stage 2. For each pair (i, j), where $j \notin A \cap B$, choose whether $\sigma_i(j)$ it will be the first appearance of the index $\sigma_i(j) \in [n]$ or if there is some $i' < i$ and j' so that $\sigma_{i'}(j') = \sigma_i(j)$. Here there are $2^{2\ell t}$ options.

Stage 3. Choose the labels $\sigma_i(j) \in [n]$ for all $j \notin A \cap B$ and pairs (i, j) which in Stage 2 we chose to be the first appearance of a label. If there are x such vertices, there are at most n^x options.

Stage 4. Choose the labels $\sigma_i(j) \in [n]$ for all the pairs (i, j), with $j \notin A \cap B$, which in Stage 2 we chose not to be the first apperance of a label. Here there are at most $x^{2\ell t - 2\ell r - x}$ options.

All together, there are at most $n^r \cdot 2^{2\ell t} \cdot n^x \cdot x^{2\ell(t-r)-x} \leqslant n^r \cdot 2^{2\ell t} \cdot n^x \cdot (2\ell t)^{2\ell(t-r)-x}$ choices for a given x. Since $4lt \ll n$, summing up over all $x \leqslant \ell(t-p-r)+p$ the total number of choices is at most $2n^r \cdot 2^{2\ell t} \cdot n^{\ell(t-p-r)+p} \cdot (2\ell t)^{\ell(t-r+p)-p}$. Putting it together,

$$\mathbb{E}\,\mathrm{Tr}(M_k^\top M_k)^\ell \leqslant 2n^r \cdot n^{\ell(t-p-r)+p} \cdot (2\ell t)^{\ell(t-r+p)-p}\,.$$

Now using Markov's inequality and standard manipulations, for any s,

$$\begin{aligned}
\mathbb{P}\{\|M_k\| \geqslant s\} &= \mathbb{P}\{\|M_k^\top M_k\|^\ell \geqslant s^{2\ell}\} \\
&\leqslant \frac{\mathbb{E}\,\|(M_k^\top M_k)^\ell\|}{s^{2\ell}} \quad \text{by Markov's} \\
&\leqslant \frac{\mathbb{E}\,\mathrm{Tr}(M_k^\top M_k)^\ell}{s^{2\ell}} \quad \text{since } \|(M_k^\top M_k)^\ell\| \leqslant \mathrm{Tr}(M_k^\top M_k)^\ell \\
&\leqslant \frac{2n^r \cdot 2^{2\ell t} \cdot n^{\ell(t-p-r)+p} \cdot (2\ell t)^{\ell(t-r+p)-p}}{s^{2\ell}}
\end{aligned}$$

Taking $\ell = (\log n)^3$ and using $p \leqslant t \leqslant O(\log n)$, there is $s = 2^t \cdot n^{(t-p-r)/2}(\log n)^{O(t-r+p)}$ so that $\mathbb{P}\{\|M_k\| \geqslant s\} \leqslant n^{-100\log n} m^{-1}$. By a union bound, $\mathbb{P}\{\|M_k\| \leqslant s \text{ for all } k\} \geqslant 1 - n^{-100\log n}$, so $\|M\| \leqslant sm$ with probability $1 - n^{-100\log n}$. Since $m \leqslant 2^{O(t)} \cdot \log(n)^{O(1)}$, this completes the proof. □

Appendix B

Omitted Details from Chapter 3

B.1 Random sparse predicates

Consider a random sparse predicate P on k variables and accepting $|P^{-1}(1)| = t$ assignments. If $t = \exp(o(k))$, we now show that P does not support a pairwise independent subgroup with high probability, as k tends to infinity. Here the randomness corresponds to choosing $P^{-1}(1)$ to be a t-sized subset of $\{0,1\}^k$ uniformly at random.

Observation B.1.1. $P^{-1}(1)$ does not contain any affine subspace of dimension 2 (over $\mathbb{F}_2$) with probability $\geqslant 1 - t^4/2^k$.

Under the condition of the observation, $P^{-1}(1)$ does not contain any pairwise independent subgroup, because any such a subgroup contains an affine subspace of dimension 2.

Proof of Observation B.1.1. Let $v_1, \ldots, v_t \in P^{-1}(1)$ be an enumeration of vectors in $P^{-1}(1)$. Note that if $P^{-1}(1)$ contains a subspace of dimension 2, then there are $1 \leqslant a < b < c \leqslant t$ such that this subspace is exactly the affine span of v_a, v_b, v_c.

For a fixed choice of the triple (a, b, c), conditioning on the event that v_a, v_b, v_c span an affine subspace of dimension 2, the remaining vector from this affine subspace also belongs to $P^{-1}(1)$ with probability at most $t/2^k$. Taking a union bound over (a, b, c) (at

most t^3 such choices), we see that $P^{-1}(1)$ contains an affine subspace with probability at most $t^4/2^k$. □

B.2 Constructing nice instances

In this section, we show the existence of *nice* instances of constraint hypergraphs and prove Theorem 3.3.4.

Lemma B.2.1. *Fix $1 > \varepsilon, \delta \geqslant 0$ and $\gamma \geqslant e^k k^2$. Then, there exists a k-uniform constraint hypergraph G with γn edges such that for $\eta = (1/\gamma^2)^{2/\delta}$, $\tau = 4\log_2(\gamma k^2)$, G:*

1. *is $(\eta n, \delta)$-expanding,*
2. *has girth $\mathfrak{g} \geqslant \log(n)/\tau$, and*

Proof. We first choose a random graph G by choosing every k uniform hyperedge, independently, with probability $p = 4\gamma \cdot k!/n^{k-1}$. Our final hypergraph will be obtained by removing hyperedges from G.

We first show that:

Claim B.2.2. For G chosen as above, with probability at least $1/3$,

1. has between $2\gamma n$ and $6\gamma n$ edges.
2. has $(\eta n, \delta)$-expansion,
3. has at most $n^{1/4}\log(n)$ cycles of length at most $\mathfrak{g}$ and

We first show that the claim above is enough to complete the proof of the lemma. We define G' to be the hypergraph obtained by removing every cycle of length at most g.By the claim above, the total number of hyperedges removed in this process, for a large enough n, is at most γn. Observe that the last property in the statement of the theorem is immediately satisfied by G'. Further, since G' is obtained only by removing hyperedges from G, G' still enjoys $(\eta n, \delta)$-expansion. Thus, G' is a constraint hypergraph that satisfies the requirements of the lemma. Finally, the total number of edges removed is sublinear in n and thus G' has at least γn edges for a large enough n.

We now move on to complete the proof of the claim above:

Proof of Claim.

1. The expected number of edges in G is given by $p \cdot \binom{n}{k} = 4\gamma n(1 - \frac{k-1}{n})^{k-1} \geqslant 4\gamma n(1 - \frac{(k-1)^2}{n})$. By an application of Chernoff bound, the probability that the number of edges does not lie in the interval $[2\gamma n, 6\gamma n]$ is at most $2e^{\frac{-\gamma n}{16}}$.

2. Next, consider any collection of s clauses and let us compute the probability that they cover at most cs variables for some $c = k - 1 - \delta$. This probability, is then upper bounded by
$$\binom{n}{cs} \cdot \left(\binom{cs}{k}\right) p^s.$$
Using that $\binom{cs}{k} \leqslant (cs)^k/k!$ and the approximation $\binom{x}{y} \leqslant \left(\frac{xe}{y}\right)^y$, we can upper bound the above expression by:
$$\left(\frac{ne}{cs}\right)^{cs} \cdot \left(\frac{e\frac{(cs)^k}{k!}}{s}\right)^s \left(\frac{\gamma \cdot k!}{n^{k-1}}\right)^s.$$
Using that $c = k - 1 - \delta$ and that $\delta < 1$ now yields an upper bound of
$$\left(\frac{s}{n}\right)^{\delta s} \cdot \left(\gamma e^k c^2\right)^s.$$

Thus, using that $\gamma > e^k k^2$ and that s satisfies $\frac{s}{n} \leqslant (1/\gamma^2)^{2/\delta}$ makes the above probability at most $(1/\gamma^2)^s$.

3. To see how to ensure that the high girth requirement, we first observe that for any integer ℓ, the expected number of cycles of length ℓ in G is at most $(dk^2)^\ell$.

   We first count the number of ways to choose a cycle of length ℓ. Recall that a cycle is given by a cyclic sequence $C_1, \ldots, C_\ell$ of hyperedges. There are $\binom{n}{k}$ ways to choose C_1, and for $2 \leqslant i < \ell$, at most $k\binom{n}{k-1}$ ways to choose the common vertex $C_{i-1} \cap C_i$ and remaining vertices for C_i, and finally at most $k^2\binom{n}{k-2}$ to choose C_ℓ that intersects both C_1 and $C_{\ell-1}$. Therefore the expected number of length-ℓ cycles is at most

$$\binom{n}{k} \cdot \left(k\binom{n}{k-1}\right)^{\ell-2} \cdot k^2\binom{n}{k-2} \cdot \left(\frac{4\gamma(k!)}{n^{k-1}}\right)^\ell \leqslant (4\gamma)^\ell k^{2\ell}. \qquad \square$$

   By an application of Markov's inequality, with probability at least 7/8 over the draw of hyperedges of G, the number of cycles of length at most $\mathfrak{g} = \frac{1}{4}\log_{\gamma k^2} n$ are at most

$$\sum_{\ell \leqslant \mathfrak{g}} (4\gamma k^2)^\ell \leqslant \mathfrak{g} n^{1/4}.$$

   By a union bound, now, all the three properties above can be ensured with probability at least 1/3.

$\square$

B.2.1 Soundness

In this section, we show that after fixing the underlying hyperedges G of an instance, with high probability over the literals on constraints, all assignments are very close to a

random assignment. Here closeness is measured with respect to the distribution $\{C(x)\}$ as one chooses a uniformly random constraint among all hyperedges of the hypergraph.

Let G be any hypergraph with m hyperedges. Let $\mathcal{I}$ be an instance with the same underlying hypergraph as G, and with the literals in all clauses be chosen uniformly at random. We have the following lemma.

Lemma B.2.3. *Suppose $m = \Omega(2^{O(k)}\varepsilon^{-2}n)$. With high probability over the choice of literals, for any assignment $x \in \{\pm 1\}^n$, the distribution $\{C(x)\}$ with C chosen uniformly at random in $\mathcal{I}$ is within ε statistical distance to the uniform distribution over $\{\pm 1\}^k$.*

Proof. Let $\mathcal{I} = (C_1, \ldots, C_m)$ be a fixed collection of literals. Let $\mu_{\mathcal{I},x}$ denotes the distribution $\{C_i(x)\}$ when i is drawn uniformly from $[m]$. For each local assignment $y \in \{\pm 1\}^k$, the probability $\mu_{\mathcal{I},x}[y]$ that a random local assignment from $\mu_{\mathcal{I},x}$ equals y is given by $\mathbb{E}_{i\in[m]}[\mathbb{1}_{C_i(x)=y}]$.

Now suppose the signs of the literals from $\mathcal{I}$ for every constraint are chosen uniformly at random, keeping the underlying subhypergraph fixed. Then $\mu_{\mathcal{I},x}[y]$ is now a random variable depending on the randomness of the literals. For each i, the indicator $\mathbb{1}_{C_i(x)=y}$ equals 1 with probability $1/2^k$, and equals 0 with the remaining probability (over the randomness of the signs of the literals on the i-th constraint), and the random variables $\mathbb{1}_{C_i(x)=y}$ are independent of each other for different i. Therefore $\mu_{\mathcal{I},x}[y]$ is the average of m independent $\{0,1\}$-indicator random variables, each being 1 with probability $1/2^k$. By Chernoff–Hoeffding bound, we have $|\mu_{\mathcal{I},x}[y] - 1/2^k| > \eta$ with probability at most $2\exp(-\eta^2 m/2^{k+1})$. By a union bound over all assignments $x \in \{\pm 1\}^n$, the maximum deviation of $\mu_{\mathcal{I},x}[y]$ from $1/2^k$ (over all x) exceeds η with probability at most

$2\exp(-\eta^2 m/2^{k+1} + n\log 2)$. Letting $\eta = \varepsilon/2^k$, we see that

$$\mathbb{P}\left[\max_x \left\{|\mu_{\mathcal{I},x}[y] - 1/2^k|\right\} \geqslant \frac{\varepsilon}{2^k}\right] \leqslant \exp(-\Omega(n))$$

as long as $m = \Omega(2^{O(k)}\varepsilon^{-2}n)$.

Now the distribution $\{C_i(x)\}$ for a random $i \in [m]$ has statistical distance at least ε implies that $|\mu_{\mathcal{I},x}[y] - 1/2^k| \geqslant \varepsilon/2^k$ for some y. By a union bound over all $y \in \{\pm 1\}^k$, the distribution $\{C_i(x)\}$ is close in statistical distance to the uniform distribution on $\{\pm 1\}^k$ except with probability $\exp(O(k) - \Omega(n))$, assuming $m = \Omega(2^{O(k)}\varepsilon^{-2}n)$. □

Bibliography

[1] *Combinatorial approaches to finding subtle signals in DNA sequences.*, volume 8, 2000. 11

[2] *How hard is it to approximate the best Nash equilibrium?*, 2009. vii

[3] *Computational Complexity and Information Asymmetry in Financial Products (Extended Abstract)*, 2010. vii, 11

[4] Probabilistic models and heuristics for the primes, 2015. Available at https://terrytao.wordpress.com/2015/01/04/254a-supplement-4-probabilistic-models-and-heuristics-for-the-primes-optional/. 220

[5] Noga Alon, Alexandr Andoni, Tali Kaufman, Kevin Matulef, Ronitt Rubinfeld, and Ning Xie. Testing k-wise and almost k-wise independence. In *STOC*, pages 496–505, 2007. vii, 11

[6] Noga Alon, Michael Krivelevich, and Benny Sudakov. Finding a large hidden clique in a random graph. In *Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, 25-27 January 1998, San Francisco, California.*, pages 594–598, 1998. 2

[7] Noga Alon, Michael Krivelevich, and Benny Sudakov. Finding a large hidden clique in a random graph. In *SODA*, pages 594–598, 1998. 11

[8] Benny Applebaum, Boaz Barak, and Avi Wigderson. Public-key cryptography from different assumptions. In *STOC*, pages 171–180, 2010. vii, 11

[9] Sanjeev Arora, Boaz Barak, and David Steurer. Subexponential algorithms for unique games and related problems. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pages 563–572. IEEE, 2010. viii

[10] Sanjeev Arora, Béla Bollobás, László Lovász, and Iannis Tourlakis. Proving integrality gaps without knowing the linear program. *Theory of Computing*, 2(1):19–51, 2006. 63

[11] Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows, geometric embeddings and graph partitioning. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 222–231. ACM, 2004. viii

[12] Per Austrin, Mark Braverman, and Eden Chlamtac. Inapproximability of np-complete variants of nash equilibrium. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques - 14th International Workshop, APPROX 2011, and 15th International Workshop, RANDOM 2011, Princeton, NJ, USA, August 17-19, 2011. Proceedings*, pages 13–25, 2011. vii

[13] Per Austrin, Mark Braverman, and Eden Chlamtac. Inapproximability of np-complete variants of nash equilibrium. *Theory of Computing*, 9:117–142, 2013. 11

[14] Per Austrin and Johan Håstad. Randomly supported independence and resistance. *SIAM J. Comput.*, 40(1):1–27, 2011. 8

[15] Per Austrin and Elchanan Mossel. Approximation resistant predicates from pairwise independence. *Computational Complexity*, 18(2):249–271, 2009. 8

[16] Eiichi Bannai and Tatsuro Ito. *Algebraic combinatorics. I*. The Benjamin/Cummings Publishing Co., Inc., Menlo Park, CA, 1984. Association schemes. 130, 151

[17] Boaz Barak. Sum of squares upper bounds, lower bounds and open questions, 2014. Lecture notes for an MIT seminar series. Available on www.boazbarak.org/sos. 61

[18] Boaz Barak. Sum of squares upper bounds, lower bounds, and open questions. *Lecture Notes*, pages http://www.boazbarak.org/sos/files/all–notes.pdf, 2014. 102, 104

[19] Boaz Barak, Fernando GSL Brandao, Aram W Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 307–326. ACM, 2012. 12, 63

[20] Boaz Barak, Siu On Chan, and Pravesh K. Kothari. Sum of squares lower bounds from pairwise independence. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 97–106, 2015. 10, 60, 100, 114, 204

[21] Boaz Barak, Sam Hopkins, Jonathan A. Kelner, Pravesh K. Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum of squares lower bound for the planted clique problem. In *Manuscript*, 2016. 14, 203

[22] Boaz Barak, Jonathan A Kelner, and David Steurer. Rounding sum-of-squares relaxations. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 31–40. ACM, 2014. 12

[23] Boaz Barak, Jonathan A. Kelner, and David Steurer. Rounding sum-of-squares relaxations. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, STOC '14, pages 31–40, New York, NY, USA, 2014. ACM. 63

[24] Boaz Barak, Jonathan A Kelner, and David Steurer. Dictionary learning and tensor decomposition via the sum-of-squares method. 2015. 12, 114

[25] Boaz Barak, Guy Kindler, and David Steurer. On the optimality of semidefinite relaxations for average-case and generalized constraint satisfaction. In *Innovations in Theoretical Computer Science, ITCS '13, Berkeley, CA, USA, January 9-12, 2013*, pages 197–214, 2013. 11

[26] Boaz Barak, Guy Kindler, and David Steurer. On the optimality of semidefinite relaxations for average-case and generalized constraint satisfaction. In *Proceedings of the 4th Conference on Innovations in Theoretical Computer Science*, ITCS '13, pages 197–214, 2013. 62

[27] Boaz Barak and Ankur Moitra. Tensor prediction, rademacher complexity and random 3-XOR. 2015. 114

[28] Boaz Barak and David Steurer. Sum-of-squares proofs and the quest toward optimal algorithms. *Proceedings of the International Congress of Mathematicians*, 2014. 61, 64, 232

[29] Boaz Barak and David Steurer. Sum-of-squares proofs and the quest toward optimal algorithms. *arXiv preprint arXiv:1404.5236*, 2014. 101, 102, 113

[30] Siavosh Bennabas, Konstantinos Georgiou, Avner Magen, and Madhur Tulsiani. SDP gaps from pairwise independence. *Theory of Computing*, 8(12):269–289, 2012. 9, 63, 64, 65, 70, 73, 74, 77

[31] Quentin Berthet and Philippe Rigollet. Complexity theoretic lower bounds for sparse principal component detection. In *Conference on Learning Theory*, pages 1046–1066, 2013. vii, 3

[32] Quentin Berthet and Philippe Rigollet. Complexity theoretic lower bounds for sparse principal component detection. In *COLT 2013 - The 26th Annual Conference on Learning Theory, June 12-14, 2013, Princeton University, NJ, USA*, pages 1046–1066, 2013. 11

[33] Aditya Bhaskara, Moses Charikar, Aravindan Vijayaraghavan, Venkatesan Guruswami, and Yuan Zhou. Polynomial integrality gaps for strong SDP relaxations of densest k-subgraph. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2012, Kyoto, Japan, January 17-19, 2012*, pages 388–405, 2012. 114

[34] Andrej Bogdanov and Luca Trevisan. On worst-case to average-case reductions for NP problems. *SIAM J. Comput.*, 36(4):1119–1159, 2006. 12

[35] Gábor Braun, Samuel Fiorini, Sebastian Pokutta, and David Steurer. Approximation

limits of linear programs (beyond hierarchies). *Math. Oper. Res.*, 40(3):756–772, 2015. 5, 18, 28

[36] S. Charles Brubaker and Santosh Vempala. Random tensors and planted cliques. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, 12th International Workshop, APPROX 2009, and 13th International Workshop, RANDOM 2009, Berkeley, CA, USA, August 21-23, 2009. Proceedings*, pages 406–419, 2009. 114

[37] Siu On Chan. Approximation resistance from pairwise independent subgroups. In *Proceedings of the Forty-fifth Annual ACM Symposium on Theory of Computing*, STOC '13, pages 447–456, New York, NY, USA, 2013. ACM. 8, 14, 62

[38] Siu On Chan, James R. Lee, Prasad Raghavendra, and David Steurer. Approximate constraint satisfaction requires large LP relaxations. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*, pages 350–359, 2013. viii, 5, 7, 18, 27

[39] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Integrality gaps for Sherali-Adams relaxations. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 283–292. ACM, 2009. 63

[40] Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM J. Comput.*, 17(2):230–261, 1988. 41, 42

[41] A Crisanti, L Leuzzi, and G Parisi. The 3-sat problem with large number of clauses in the ∞-replica symmetry breaking scheme. *Journal of Physics A: Mathematical and General*, 35(3):481, 2002. 3

[42] Wenceslas Fernandez de la Vega and Claire Kenyon-Mathieu. Linear programming relaxations of maxcut. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 53–61. Society for Industrial and Applied Mathematics, 2007. 63

[43] Yash Deshpande and Andrea Montanari. Improved sum-of-squares lower bounds for hidden clique and hidden submatrix problems. *COLT*, 2015. 12, 13, 98, 99, 100, 101, 102, 107, 108, 109, 110, 111, 115, 117, 119, 139, 143, 150, 151, 152, 209, 210, 212, 258

[44] Uriel Feige. A threshold of ln n for approximating set cover. *J. ACM*, 45(4):634–652, 1998. 2

[45] Uriel Feige and Robert Krauthgamer. Finding and certifying a large hidden clique in a semirandom graph. *Random Struct. Algorithms*, 16(2):195–208, 2000. 114

[46] Uriel Feige and Robert Krauthgamer. The probable value of the lovász–schrijver relaxations for maximum independent set. *SIAM J. Comput.*, 32(2):345–370, 2003. 12, 13, 99, 100, 103, 114

[47] Uriel Feige and Robert Krauthgamer. The probable value of the lovász–schrijver relaxations for maximum independent set. *SIAM J. Comput.*, 32(2):345–370, 2003. 209, 210, 211, 212

[48] Joan Feigenbaum and Lance Fortnow. Random-self-reducibility of complete sets. *SIAM J. Comput.*, 22(5):994–1005, 1993. 12

[49] Vitaly Feldman, Elena Grigorescu, Lev Reyzin, Santosh Vempala, and Ying Xiao. Statistical algorithms and a lower bound for planted cliques. In *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 655–664, 2013. 114

[50] Samuel Fiorini, Serge Massar, Sebastian Pokutta, Hans Raj Tiwary, and Ronald de Wolf. Exponential lower bounds for polytopes in combinatorial optimization. *J. ACM*, 62(2):17, 2015. 19

[51] A. Frieze and R. Kannan. The regularity lemma and approximation schemes for dense problems. In *Foundations of Computer Science, 1996. Proceedings., 37th Annual Symposium on*, pages 12–20, Oct 1996. 220

[52] Alan M. Frieze and Ravi Kannan. A new approach to the planted clique problem. In *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2008, December 9-11, 2008, Bangalore, India*, pages 187–198, 2008. 114

[53] Rong Ge and Tengyu Ma. Decomposing overcomplete 3rd order tensors using sum-of-squares algorithms. *In Proc. APPROX/RANDOM*, abs/1504.05287, 2015. 114

[54] Michel X Goemans and David P Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42(6):1115–1145, 1995. viii, 2, 4, 63

[55] Oded Goldreich. Average case complexity, revisited. In *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation - In Collaboration with Lidor Avigad, Mihir Bellare, Zvika Brakerski, Shafi Goldwasser, Shai Halevi, Tali Kaufman, Leonid Levin, Noam Nisan, Dana Ron, Madhu Sudan, Luca Trevisan, Salil Vadhan, Avi Wigderson, David Zuckerman*, pages 422–450. 2011. 2

[56] Mika Göös, Shachar Lovett, Raghu Meka, Thomas Watson, and David Zuckerman. Rectangles are nonnegative juntas. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 257–266, 2015. 22, 41

[57] Andrew Granville. Harald cramér and the distribution of prime numbers. *Scandinavian Actuarial Journal*, 1995(1):12–28, 1995. 220

[58] Dima Grigoriev. Complexity of positivstellensatz proofs for the knapsack. *Computational Complexity*, 10(2):139–154, 2001. 9, 63

[59] Dima Grigoriev. Complexity of positivstellensatz proofs for the knapsack. *Computational Complexity*, 10(2):139–154, 2001. 14, 62, 100, 110, 113, 204, 205, 209

[60] Dima Grigoriev. Linear lower bound on degrees of positivstellensatz calculus proofs for the parity. *Theor. Comput. Sci.*, 259(1-2):613–622, 2001. 113

[61] Bruce E. Hajek, Yihong Wu, and Jiaming Xu. Computational lower bounds for community detection on random graphs. In *Proceedings of The 28th Conference on Learning Theory, COLT 2015, Paris, France, July 3-6, 2015*, pages 899–928, 2015. 11

[62] Johan Håstad. Clique is hard to approximate within $n^{1-epsilon}$. *Electronic Colloquium on Computational Complexity (ECCC)*, 4(38), 1997. 2

[63] Johan Håstad. Some optimal inapproximability results. *J. ACM*, 48(4):798–859, 2001. 2, 4

[64] Elad Hazan and Robert Krauthgamer. How hard is it to approximate the best nash equilibrium? *SIAM J. Comput.*, 40(1):79–91, 2011. 11

[65] Samuel B. Hopkins, Pravesh Kothari, Aaron Henry Potechin, Prasad Raghavendra, and Tselil Schramm. On the integrality gap of degree-4 sum of squares for planted clique. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 1079–1095, 2016. 213, 258

[66] Samuel B. Hopkins, Pravesh K. Kothari, and Aaron Potechin. Sos and planted clique: Tight analysis of MPW moments at all degrees and an optimal lower bound at degree four. *CoRR*, abs/1507.05230, 2015. 13, 98, 212, 213, 221, 222, 223, 258

[67] Samuel B. Hopkins, Jonathan Shi, and David Steurer. Tensor principle component analysis via sum of squares proofs. *In Proc. COLT*, 2015. 114

[68] Russell Impagliazzo, Ramamohan Paturi, and Francis Zane. Which problems have strongly exponential complexity? *J. Comput. Syst. Sci.*, 63(4):512–530, 2001. 1

[69] Svante Janson, Tomasz Luczak, and Andrzej Rucinski. *Random graphs*, volume 45. John Wiley & Sons, 2011. 186

[70] Hamid Javadi and Andrea Montanari. The hidden subgraph problem. *arXiv preprint arXiv:1511.05254*, 2015. 11

[71] E. T. Jaynes. Information theory and statistical mechanics. ii. *Phys. Rev.*, 108:171–190, Oct 1957. 214

[72] Edwin T Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957. 214

[73] Mark Jerrum. Large cliques elude the metropolis process. *Random Struct. Algorithms*, 3(4):347–360, 1992. 11, 114

[74] Ari Juels and Marcus Peinado. Hiding cliques for cryptographic security. *Des. Codes Cryptography*, 20(3):269–280, July 2000. 11

[75] Richard M. Karp. Probabilistic analysis of some combinatorial search problems. *Algorithms and Complexity: New Directions and Recent Results*, 1976. 11

[76] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings on 34th Annual ACM Symposium on Theory of Computing, May 19-21, 2002, Montréal, Québec, Canada*, pages 767–775, 2002. viii, 8

[77] Pascal Koiran and Anastasios Zouzias. Hidden cliques and the certification of the restricted isometry property. *IEEE Trans. Information Theory*, 60(8):4999–5006, 2014. 11

[78] Pravesh K. Kothari, Raghu Meka, and Prasad Raghvendra. Approximating rectangles by juntas and a sub-exponential lower bound on linear programs for max-csp. In *Manuscript*, 2016. 6, 16

[79] Ludek Kucera. Expected complexity of graph partitioning problems. *Discrete Applied Mathematics*, 57(2-3):193–212, 1995. 11

[80] Jean B. Lasserre. An explicit exact sdp relaxation for nonlinear 0-1 programs. In *IPCO*, pages 293–303, 2001. vii, 9, 12, 62

[81] James R Lee, Prasad Raghavendra, and David Steurer. Lower bounds on the size of semidefinite programming relaxations. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*. ACM, 2015. 27

[82] Tengyu Ma and Avi Wigderson. Sum of squares lower bounds for sparse pca. *Preprint*, 2015. 114

[83] Tengyu Ma and Avi Wigderson. Sum-of-squares lower bounds for sparse pca. In *Advances in Neural Information Processing Systems*, pages 1603–1611, 2015. 211

[84] Dhruv Medarametla and Aaron Potechin. Bounds on the norms of uniform low degree graph matrices. *Preprint*. 229, 258, 284

[85] Raghu Meka, Aaron Potechin, and Avi Wigderson. Sum-of-squares lower bounds for planted clique. pages 87–96, 2015. 12, 13, 98, 99, 100, 101, 102, 103, 111, 115, 117, 118, 119, 120, 121, 130, 142, 143, 150, 151, 155, 180, 181, 185, 186, 209, 210, 212, 221

[86] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: Simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002. 11

[87] Andrea Montanari, Daniel Reichman, and Ofer Zeitouni. On the limitation of spectral methods: From the gaussian hidden clique problem to rank one perturbations of gaussian tensors. *Arxiv:1411.6149*, 2014. 114

[88] Yurii Nesterov. Squared functional systems and optimization problems. *High performance optimization*, 13:405–440, 2000. 9

[89] Ryan O'Donnell and David Witmer. Goldreich's PRG: Evidence for near-optimal polynomial stretch. In *IEEE Conference on Computational Complexity*, pages 1–12, 2014. 9, 64

[90] Pablo A. Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. PhD thesis, California Institute of Technology, May 2000. vii, 9, 12, 62

[91] Kanstantsin Pashkovich. Tight lower bounds on the sizes of symmetric extensions of permutahedra and similar results. *Math. Oper. Res.*, 39(4):1330–1339, 2014. 28

[92] Prasad Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, STOC '08, pages 245–254, New York, NY, USA, 2008. ACM. 9

[93] Prasad Raghavendra and Tselil Schramm. Tight lower bounds for planted clique in the degree-4 sos program. *Preprint*, 2015. 13, 101

[94] Andrzej Ruciński. When are small subgraphs of a random graph normally distributed? *Probability Theory and Related Fields*, 78(1):1–10, 1988. 186

[95] Grant Schoenebeck. Linear level lasserre lower bounds for certain k-csps. In *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*, pages 593–602, 2008. 9, 62, 100, 113

[96] Grant Schoenebeck. Linear level lasserre lower bounds for certain k-csps. In *FOCS*, pages 593–602, 2008. 14, 204, 205

[97] Grant Schoenebeck, Luca Trevisan, and Madhur Tulsiani. Tight integrality gaps for Lovász-Schrijver LP relaxations of vertex cover and max cut. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 302–310. ACM, 2007. 63

[98] Jean-Pierre Serre. *Linear representations of finite groups*, volume 42. Springer Science & Business Media, 2012. 130

[99] Alexander A. Sherstov. The pattern matrix method. *SIAM J. Comput.*, 40(6):1969–2000, 2011. 19

[100] N. Z. Shor. Class of global minimum bounds of polynomial functions. *Cybernetics*, 23(6):731–734, 1987. (Russian orig.: Kibernetika, No. 6, (1987), 9–11). vii, 9, 12

[101] Endre Szemerédi. Regular partitions of graphs. *Problèmes combinatoires et théorie des graphes*, 1978. 220

[102] Terence Tao. The dichotomy between structure and randomness, arithmetic progressions, and the primes. *arXiv preprint math/0512114*, 2005. 219

[103] Madhur Tulsiani. CSP gaps and reductions in the Lasserre hierarchy. In *STOC*, pages 303–312, 2009. 9, 62

[104] Madhur Tulsiani. CSP gaps and reductions in the lasserre hierarchy. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 2, 2009*, pages 303–312, 2009. 114

[105] Madhur Tulsiani and Pratik Worah. LS+ lower bounds from pairwise independence. In *Proceedings of the 28th Conference on Computational Complexity, CCC*, pages 121–132, 2013. 9, 64, 73, 74

[106] Van H Vu. A large deviation result on the number of small subgraphs of a random graph. *Combinatorics, Probability & Computing*, 10(1):79–94, 2001. 186

[107] Mihalis Yannakakis. Expressing combinatorial optimization problems by linear programs (extended abstract). In *Proceedings of the 20th Annual ACM Symposium on Theory of Computing, May 2-4, 1988, Chicago, Illinois, USA*, pages 223–228, 1988. viii, 5, 18, 28

Vita

Pravesh Kumar Kothari, the son of Omprakash S. Kothari and Sulabha Kothari was born in Yavatmal, Maharashtra, India on 3rd March 1989. After completing his high school education from Hyderabad, India, he attended Indian Institute of Technology, Kanpur and received his Bachelor of Technology degree in Electrical Engineering in 2010. He joined the graduate school at the University of Texas at Austin in January 2011.

Permanent address: kotpravesh@gmail.com

This dissertation was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.